CHAPTER 5

# MINIMUM DISTANCE ESTIMATORS

## 5.1. INTRODUCTION

The practice of obtaining estimators of parameters by minimizing a certain distance between some functions of observations and parameters has long been present in statistics. The classical examples of this method are the Least Square and the minimum Chi Square estimators.

The minimum distance estimation (m.d.e.) method, where one obtains an estimator of a parameter by minimizing some distance between the empirical d.f. and the modeled d.f., was elevated to a general method of estimation by Wolfowitz (1953, 1954, 1957). In these papers he demonstrated that compared to the maximum likelihood estimation method, the m.d.e. method yielded consistent estimators rather cheaply in several problems of varied levels of difficulty.

This methodology saw increasing research activity from the mid–seventy's when many authors demonstrated various robustness properties of certain m.d. estimators. Beran (1977) showed that in the i.i.d. setup the minimum Hellinger distance estimators, obtained by minimizing the Hellinger distance between the modeled parametric density and an empirical density estimate, are asymptotically efficient at the true model and robust against small departures from the model, where the smallness is being measured in terms of the Hellinger metric. Beran (1978) demonstrated the powerfulness of minimum Hellinger distance estimators in the one sample location model by showing that the estimators obtained by minimizing the Hellinger distance between an estimator of the density of the residual and an estimator of the density of the negative residual are qualitatively robust and adaptive for all those symmetric error distributions that have finite Fisher information.

Parr and Schucany (1979) empirically demonstrated that in certain location models several minimum distance estimators (where several comes from the type of distances chosen) are robust. Millar (1981, 1982, 1984) proved local asymptotic minimaxity of a fairly large class of m.d. estimators, using Cramer-Von Mises type distance, in the i.i.d. setup. Donoho and Liu (1988 a, b) demonstrated certain further finite sample robustness properties of a large class of m.d. estimators and certain additional advantages of using Cramer-Von Mises and Hellinger distances. All of these authors restrict their attention to the one sample setup or to the two sample location model. See Parr (1981) for additional bibliography on m.d.e. through 1980.

Little was known till the early 1980's about how to extend the above methodology to one of the most applied models, v.i.z., the multiple linear regression model (1.1.1). Given the above optimality properties in the one- and two- sample location models, it became even more desirable to extend this methodology to this model. Only after realizing that one should use the weighted, rather than the ordinary, empiricals of the residuals to define m.d. estimators was it possible to extend this methodology satisfactorily to the model (1.1.1).

The main focus of this chapter is the m.d. estimators of $\boldsymbol{\beta}$ obtained by minimizing the Cramer-Von Mises type distances involving various w.e.p.'s. Some m.d. estimators involving the supremum distance are also discussed. Most of the estimators provide appropriate extensions of their counterparts in the one- and two- sample location models.

Section 5.2 contains definitions of several m.d. estimators. Their finite sample properties and asymptotic distributions are discussed in Sections 5.3, 5.5, respectively. Section 5.4 discusses an asymptotic theory about general minimum dispersion estimators that is of broad and independent interest. It is a self contained section. Asymptotic relative efficiency and qualitative robustness of some of the m.d. estimators of Section 5.2 are discussed in Section 5.6. Some of the proposed m.d. functionals are Hellinger differentiable in the sense of Beran (1982) as is shown in Section 5.6. Consequently they are locally asymptotically minimax (l.a.m.) in the sense of Hájek – Le Cam.

## 5.2. DEFINITIONS OF M.D. ESTIMATORS

To motivate the following definitions of m.d. estimators of $\boldsymbol{\beta}$ of (1.1.1), first consider the one sample location model where $Y_1 - \theta, \ldots., Y_n - \theta$ are i.i.d. F, F a *known* d.f.. Let

$$F_n(y) := n^{-1} \sum_{i=1}^{n} I(Y_i \le y), \qquad y \in \mathbb{R}. \tag{1}$$

If $\theta$ is true then $EF_n(y + \theta) = F(y)$, $\forall\, y \in \mathbb{R}$. This motivates one to define m.d. estimator $\hat{\theta}$ of $\theta$ by the relation

$$\hat{\theta} = \operatorname{argmin}\{T(t);\ t \in \mathbb{R}\} \tag{2}$$

where, for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$,

$$T(t) := n \int [F_n(y + t) - F(y)]^2\, dG(y), \qquad t \in \mathbb{R}. \tag{3}$$

Observe that (2) and (3) actually define a class of estimators $\hat{\theta}$, one corresponding to each $G$.

Now suppose that in (1.1.1) we model the d.f. of $e_{ni}$ to be a *known* d.f. $H_{ni}$, which may be different from the actual d.f. $F_{ni}$, $1 \le i \le n$. How should one define a m.d. estimator of $\boldsymbol{\beta}$? Any definition should reduce to $\hat{\theta}$ when (1.1.1) is reduced to the one sample location model. One possible extension is to define

$$\hat{\boldsymbol{\beta}}_1 = \operatorname{argmin}\{K_1(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}, \tag{4}$$

where

(5) $$K_1(\mathbf{t}) = n^{-1}\int [\sum_{i=1}^{n}\{I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - H_{ni}(y)\}]^2\, dG(y), \quad \mathbf{t}\in\mathbb{R}^p.$$

If in (1.1.1) we take $p = 1$, $x_{ni1} \equiv 1$ and $H_{ni} \equiv F$ then clearly it reduces to the one sample location model and $\hat{\beta}_1$ coincides with $\hat{\theta}$ of (2). But this is also true for the estimator $\hat{\beta}_{\mathbf{X}}$ defined as follows. Recall the definition of $\{V_j\}$ from (1.2.1). Define, for $y \in \mathbb{R}$, $\mathbf{t} \in \mathbb{R}^p$, $1 \le j \le p$,

(6) $$\mathcal{Z}_j(y, \mathbf{t}) := V_j(y, \mathbf{t}) - \sum_{i=1}^{n} x_{nij}\, H_{ni}(y).$$

Let

(7) $$K_{\mathbf{X}}(\mathbf{t}) := \int \mathcal{Z}'(y, \mathbf{t})(X'X)^{-1}\mathcal{Z}(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\mathcal{Z}' := (\mathcal{Z}_1, \ldots., \mathcal{Z}_p)$ and define,

(8) $$\hat{\beta}_{\mathbf{X}} = \operatorname{argmin}\{K_{\mathbf{X}}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

Which of the two estimators is the right extension of $\hat{\theta}$ ? Since $\{V_j, 1 \le j \le p\}$ summarize the data in (1.1.1) with probability one under the continuity assumption of $\{e_{ni}, 1 \le i \le n\}$, $\hat{\beta}_{\mathbf{X}}$ should be considered the right extension of $\hat{\theta}$. In Section 5.6 we shall see that $\hat{\beta}_{\mathbf{X}}$ is asymptotically efficient among a class of estimators $\{\hat{\beta}_{\mathbf{D}}\}$ defined as follows.

Let $\mathbf{D} = ((d_{nij}))$, $1 \le i \le n$, $1 \le j \le p$, be an $n{\times}p$ real matrix,

(9) $$V_{jd}(y, \mathbf{t}) := \sum_{i=1}^{n} d_{nij}\, I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \qquad y \in \mathbb{R},\ 1 \le j \le p,$$

and

(10) $$K_{\mathbf{D}}(\mathbf{t}) := \sum_{j=1}^{p}\int [V_{jd}(y, \mathbf{t}) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]^2\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p.$$

Define

(11) $$\hat{\beta}_{\mathbf{D}} = \operatorname{argmin}\{K_{\mathbf{D}}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

If $\mathbf{D} = n^{-1/2}[1,0, \ldots, 0]_{n\times p}$ then $\hat{\beta}_{\mathbf{D}} = \hat{\beta}_1$ and if $\mathbf{D} = \mathbf{X}A$ then $\hat{\beta}_{\mathbf{D}} = \hat{\beta}_{\mathbf{X}}$, where $\mathbf{A}$ is as in (2.3.32). The above mentioned optimality of $\hat{\beta}_{\mathbf{X}}$ is stated and proved in Theorem 5.6a.1.

Another way to define m.d. estimators in the case the *modeled error d.f.'s are known* is as follows. Let

$$(12)\qquad M(s,y,\mathbf{t}) := n^{-1/2} \sum_{i=1}^{ns} \{I(Y_{ni} \le y) - H_{ni}(y - \mathbf{x}_{ni}'\mathbf{t})\}, \quad s \in [0, 1],\ y \in \mathbb{R},$$

$$(13)\qquad Q(\mathbf{t}) := \int_0^1 \int \{M(s,y,\mathbf{t})\}^2\, dG(y)\, dL(s), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where L is a d.f. on $[0, 1]$. Define

$$(14)\qquad \bar{\beta} = \operatorname{argmin}\{Q(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

The estimator $\bar{\beta}$ with $L(s) \equiv s$ is essentially Millar's (1982) proposal.

Now suppose $\{H_{ni}\}$ are *unknown*. How should one define m.d. estimators of $\beta$ in this case? Again, let us examine the one sample location model. In this case $\theta$ can not be identified unless the errors are symmetric about 0. Suppose that is the case. Then the r.v.'s $\{Y_i - \theta,\ 1 \le i \le n\}$ have the same distribution as $\{-Y_i + \theta,\ 1 \le i \le n\}$. A m.d. estimator $\theta^+$ of $\theta$ is thus defined by the relation

$$(15)\qquad \theta^+ = \operatorname{argmin}\{T^+(t),\ t\in\mathbb{R}\}$$

where

$$(16)\qquad T^+(t) := n^{-1} \int \Big[\sum_{i=1}^{n} \{I(Y_i \le y + t) - I(-Y_i < y - t)\}\Big]^2 dG(y).$$

An extension of $\theta^+$ to the model (1.1.1) is $\beta_{\mathbf{X}}^+$ defined by the relation

$$(17)\qquad \beta_{\mathbf{X}}^+ = \operatorname{argmin}\{K_{\mathbf{X}}^+(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^p\}$$

where, for $\mathbf{t} \in \mathbb{R}^p$,

$$(18)\qquad K_{\mathbf{X}}^+(\mathbf{t}) := \int \mathbf{V}^+(y, \mathbf{t})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}^+(y, \mathbf{t})\, dG(y), \qquad \mathbf{V}^{+\prime} = (V_1^+, \ldots, V_p^+),$$

$$V_j^+(y, \mathbf{t}) := \sum_{i=1}^{n} x_{nij}\{I(Y_{ni} \le y+\mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y-\mathbf{x}_{ni}'\mathbf{t})\},\quad y\in\mathbb{R},\ 1\le j\le p.$$

More generally, a class of m.d. estimators of $\beta$ can be defined as follows. Let $\mathbf{D}$ be as before. Define, for $y\in\mathbb{R}$, $1 \le j \le p$,,

$$(19a)\qquad Y_j^+(y, \mathbf{t}) := \sum_{i=1}^{n} d_{nij}\, \{I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})\}.$$

Let $\mathbf{Y}_{\mathbf{D}}^{+'} = (Y_1^+, \ldots, Y_p^+)$ and define

$$\text{(19b)} \qquad K_{\mathbf{D}}^{+}(\mathbf{t}) := \int \mathbf{Y}_{\mathbf{D}}^{+'}(y, \mathbf{t})\, \mathbf{Y}_{\mathbf{D}}^{+}(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p.$$

and $\boldsymbol{\beta}_{\mathbf{D}}^{+}$ by the relation

$$\text{(20)} \qquad \boldsymbol{\beta}_{\mathbf{D}}^{+} = \operatorname{argmin}\{K_{\mathbf{D}}^{+}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

Note that $\boldsymbol{\beta}_{\mathbf{X}}^{+}$ is $\boldsymbol{\beta}_{\mathbf{D}}^{+}$ with $\mathbf{D} = \mathbf{XA}$.

Next, suppose that the errors in (1.1.1) are modeled to be i.i.d., i.e., $H_{ni} \equiv F$ and $F$ is *unknown* and *not necessarily symmetric*. Here, of course, the location parameter can not be estimated. However, the regression parameter vector $\boldsymbol{\beta}$ can be estimated provided the rank of $\mathbf{X}_c$ is $p$, where $\mathbf{X}_c$ is defined at (4.3.11). In this case a class of m.d. estimators of $\boldsymbol{\beta}$ is defined by $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ of (11) provided we assume that

$$\text{(21)} \qquad \sum_{i=1}^{n} d_{nij} = 0, \qquad 1 \le j \le p.$$

A member of this class that is of interest is $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ with $\mathbf{D} = \mathbf{X}_c\mathbf{A}_1$, $\mathbf{A}_1$ as in (4.3.11).

Another way to define m.d. estimators here is via the ranks. With $R_{it}$ as in (3.1.1), let

$$\text{(22)} \qquad T_{jd}(s, \mathbf{t}) := \sum_{i=1}^{n} d_{nij}\, I(R_{it} \le ns), \qquad s\in[0, 1],\ 1 \le j \le p,$$

$$K_{\mathbf{D}}^{*}(\mathbf{t}) := \int \mathbf{T}_{\mathbf{D}}'(s, \mathbf{t})\, \mathbf{T}_{\mathbf{D}}(s, \mathbf{t})\, dL(s), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\mathbf{T}_{\mathbf{D}}' = (T_1, \ldots, T_p)$ and $L$ is a d.f. on $[0, 1]$. Assume that $\mathbf{D}$ satisfies (21). Define

$$\text{(23)} \qquad \boldsymbol{\beta}_{\mathbf{D}}^{*} = \operatorname{argmin}\{K_{\mathbf{D}}^{*}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

Observe that $\{\hat{\boldsymbol{\beta}}_{\mathbf{D}}\}$, $\{\boldsymbol{\beta}_{\mathbf{D}}^{+}\}$ and $\{\bar{\boldsymbol{\beta}}\}$ are not scale invariant in the sense of (4.3.2). One way to make them so is to modify their definitions as follows. Define

$$\text{(24)} \qquad K_{\mathbf{D}}(a, \mathbf{t}) := \sum_{j=1}^{p} \int [V_{jd}(ay, \mathbf{t}) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]^2\, dG(y),$$

$$K_{\mathbf{D}}^{+}(a, \mathbf{t}) := \int \mathbf{Y}_{\mathbf{D}}^{+'}(ay, \mathbf{t}) \mathbf{Y}_{\mathbf{D}}^{+}(ay, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,\ a \ge 0.$$

Now, scale invariant analogues of $\hat{\beta}_D$ and $\beta_D^+$ are defined as

$$(25) \qquad \hat{\beta}_D^o := \text{argmin}\ \{K_D(s, t),\ t\in\mathbb{R}^p\}, \quad \beta_D^{+o} := \text{argmin}\ \{K_D^+(s, t),\ t\in\mathbb{R}^p\},$$

where $s$ is a scale estimator satisfying (4.3.3) and (4.3.4). One can modify $\{\bar{\beta}\}$ in a similar fashion to make it scale invariant. The class of estimators $\{\beta_D^*\}$ is scale invariant because the ranks are.

Now we define a m.d. estimator based on the *supremum distance* in the case the errors are correctly modeled to be i.i.d. F, F an arbitrary d.f. Here we shall *restrict* ourselves only to the *case of* $p = 1$. Define

$$(26) \qquad V_c(y, t) := \sum_{i=1}^n (x_i - \bar{x})\, I(Y_i \le y + tx_i), \qquad t, y \in \mathbb{R},$$

$$\mathcal{D}_n^+(t) := \sup\ \{V_c(y, t);\ y\in\mathbb{R}\},$$

$$\mathcal{D}_n^-(t) := -\inf\ \{V_c(y, t);\ y\in\mathbb{R}\},$$

$$\mathcal{D}_n(t) := \max\ \{\mathcal{D}_n^+(t), \mathcal{D}_n^-(t)\} = \sup\{|V_c(y, t)|;\ y\in\mathbb{R}\}, \quad t \in \mathbb{R}.$$

Finally, define the m.d. estimator

$$(27) \qquad \hat{\beta}_s := \text{argmin}\{\mathcal{D}_n(t);\ t\in\mathbb{R}\}.$$

Section 5.3 discusses some computational aspects including the existence and some finite sample properties of the above estimators. Section 5.5 proves the uniform asymptotic quadraticity of $K_D$, $K_D^+$, $K_D^*$ and $Q$ as processes in $t$. These results are used in Section 5.6 to study the asymptotic distributions and robustness of the above defined estimators.

## 5.3 FINITE SAMPLE PROPERTIES AND EXISTENCE

The purpose here is to discuss some computational aspects, the existence and the finite sample properties of the *four* classes of estimators introduced in the previous section. To facilitate this the dependence of these estimators and their defining statistics on the weight matrix D will not be exhibited in this section.

We first turn to *some computational aspects* of these estimators. To begin with, suppose that $p = 1$ and $G(y) = y$ in (5.2.10) and (5.2.11). Write $\hat{\beta}$, $x_i$, $d_i$ for $\hat{\beta}$, $x_{i1}$, $d_{i1}$, respectively, $1 \le i \le n$. Then

(1) $$K(t) = \int [\Sigma_i d_i\{I(Y_i \le y + x_i t) - H_i(y)\}]^2 \, dy$$
$$= \Sigma_i \Sigma_j d_i d_j \int \{I(Y_i \le y+x_i t) - H_i(y)\}\{I(Y_j \le y+x_j t) - H_j(y)\} dy.$$

No further simplification of this occurs except for some special cases. One of them is the case of the one sample location model where $x_i \equiv 1$ and $H_i \equiv F$, in which case

$$K(t) = \int [\Sigma_i d_i\{I(Y_i \le y) - F(y - t)\}]^2 \, dy.$$

Differentiating under the integral sign w.r.t. t (which can be justified under the sole assumption: $F$ has a density $f$ w.r.t $\lambda$) one obtains

$$\dot{K}(t) = 2\int [\Sigma_i d_i\{I(Y_i \le y + t) - F(y)\} \, dF(y)$$
$$= -2 \, \Sigma_i d_i\{F(Y_i - t) - 1/2\}.$$

Upon taking $d_i \equiv n^{-1/2}$ one sees that in the one sample location model $\hat{\theta}$ of (5.2.2) corresponding to $G(y) = y$ is given as a solution of

(2) $$\Sigma_i F(Y_i - \hat{\theta}) = n/2.$$

Note that this $\hat{\theta}$ is precisely the m.l.e. of $\theta$ when $F(x) \equiv \{1 + \exp(-x)\}^{-1}$, i.e., when the errors have logistic distribution!

Another simplification of (1) occurs when we assume $\Sigma_i d_i = 0$ and $H_i \equiv F$. Fix a $t \in \mathbb{R}$ and let $c := \max\{Y_i - x_i t;\ 1 \le i \le n\}$. Then

(3) $$K(t) = \int [\Sigma_i d_i I(Y_i \le y + x_i t)]^2 \, dy$$
$$= \Sigma_i \Sigma_j d_i d_j \int I[\max(Y_j - x_j t, Y_i - x_i t) \le y < c] \, dy$$
$$= -\Sigma_i \Sigma_j d_i d_j \max(Y_j - x_j t, Y_i - x_i t).$$

Using the relationship

(4) $$2 \max(a, b) = a + b + |a-b|, \qquad a, b \in \mathbb{R},$$

and the assumption $\Sigma_i d_i = 0$, one obtains

(5) $$K(t) = -2 \sum_{1 \le i < j \le n} d_i d_j |Y_j - Y_i - (x_j - x_i)t|$$

If $d_i = x_i - \bar{x}$ in (5), then the corresponding $\hat{\beta}$ is asymptotically equivalent to the Wilcoxon type R-estimator of $\beta$ as was shown by Williamson (1979). The result will also follow from the general asymptotic theory of Sections 5 and 6.

If $d_i = x_i - \bar{x}$, $1 \le i \le n$, and $x_i = 0$, $1 \le i \le r$; $x_i = 1$, $r+1 \le i \le n$ then (1.1.1) becomes the two sample location model and

$$K(t) = -2 \sum_{i=1}^{r} \sum_{j=r+1}^{n} |Y_j - Y_i - t| + \text{a r.v. constant in } t.$$

Consequently here $\hat{\beta} = \text{med}\{|Y_j - Y_i|,\ r+1 \le j \le n,\ 1 \le i \le r\}$, the usual Hodges–Lehmann estimator. The fact that in the two sample location model the Cramer–Von Mises type m.d. estimator of the location parameter is the Hodges–Lehmann estimator was first noted by Fine (1966).

Note that a relation like (5) is true for general p and G. That is, suppose that $p \ge 1$, $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$ and (5.2.21) holds, then $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$K(\mathbf{t}) = -2 \sum_{j=1}^{p} \sum\sum_{1 \le i < k \le n} d_{ij} d_{kj}\, |G((Y_k - \mathbf{x}_k'\mathbf{t})_-) - G((Y_i - \mathbf{x}_i'\mathbf{t})_-)|. \tag{6}$$

To prove this proceed as in (3) to conclude first that

$$K(\mathbf{t}) = -2 \sum_{j=1}^{p} \sum\sum_{1 \le i < k \le n} d_{ij} d_{kj}\, G(\max(Y_k - \mathbf{x}_k'\mathbf{t},\ Y_i - \mathbf{x}_i'\mathbf{t})_-)$$

Now use the fact that $G((a \vee b)_-) = G(a_-) \vee G(b_-)$, (5.2.21) and (4) to obtain (6). Clearly, formula (6) can be used to compute $\hat{\beta}$ in general.

Next *consider* $K^+$. To simplify the exposition, fix a $\mathbf{t} \in \mathbb{R}^p$ and let $r_i := Y_i - \mathbf{x}_i'\mathbf{t}$, $1 \le i \le n$; $b := \max\{r_i, -r_i;\ 1 \le i \le n\}$. Then from (5.2.19) we obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(r_i \le y) - I(-r_i < y)\}]^2\, dG(y).$$

Observe that the integrand is zero for $y > b$. Now expand the quadratic and integrate term by term, noting that G may have jumps, to obtain

$$\begin{aligned} K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_i \sum_k d_{ij} d_{kj} \big\{ & 2G(r_i \vee -r_k)\} - 2J(r_i) \\ & - G((r_i \vee r_k)_-) - G(-r_i \vee -r_k) \big\}, \end{aligned}$$

where $J(y) := G(y) - G(y_-)$, the jump in G at $y \in \mathbb{R}$. Once again use the fact that $G(a \vee b) = G(a) \vee G(b)$, (4), the invariance of the double sum under permutation and the definition of $\{r_i\}$ to conclude that

$$(7) \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_i \sum_k d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| - J(Y_i - \mathbf{x}_i'\mathbf{t}) - \tfrac{1}{2}\Big\{|G((Y_i - \mathbf{x}_i'\mathbf{t})-) - G((Y_k - \mathbf{x}_k'\mathbf{t})-)| + |G(-Y_i + \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})|\Big\}\Big].$$

Before proceeding further it is convenient to recall at this time the definition of symmetry for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$.

**Definition 5.3.1.** An arbitrary $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$, inducing a $\sigma$–finite measure on the Borel line $(\mathbb{R}, \mathscr{B})$, is said to be *symmetric* around 0 if

$$(8) \qquad |G(y) - G(x)| = |G(-x-) - G(-y-)|, \qquad \forall\ x, y \in \mathbb{R}.$$

or

$$(9) \qquad dG(y) = -\, dG(-y), \qquad \forall\ y \in \mathbb{R}.$$

If $G$ is continuous then (8) is equivalent to

$$(10) \qquad |G(y) - G(x)| = |G(-x) - G(-y)|, \qquad \forall\ x, y \in \mathbb{R}.$$

Conversely, if (10) holds then $G$ is symmetric around 0 and continuous.

Now suppose that $G$ satisfies (8). Then (7) simplifies to

$$(7') \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \Sigma_i \Sigma_k\, d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| - J(Y_i - \mathbf{x}_i'\mathbf{t}) - |G(-Y_i + \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})|\Big].$$

And if $G$ satisfies (10) then we obtain the relatively simpler expression

$$(7^*) \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \Sigma_i \Sigma_k\, d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| - |G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(Y_k - \mathbf{x}_k'\mathbf{t})|\Big].$$

Upon specializing (7*) to the case $G(y) = y$, $p = 1$, $d_i \equiv n^{-1/2}$ and $x_i \equiv 1$ we obtain

$$K^+(t) = n^{-1}\, \Sigma_i \Sigma_k\, \{|Y_i + Y_k - 2t| - |Y_i - Y_k|\}$$

and the corresponding minimizer is the well celebrated median of the pairwise means $\{(Y_i+Y_j)/2;\ 1 \le i \le j \le n\}$.

Suppose we specialize (1.1.1) to a completely randomized design with p treatments, i.e., take

$$\begin{aligned} x_{ij} &= 1, && m_{j-1} + 1 \le i \le m_j, \\ &= 0, && \text{otherwise,} \end{aligned}$$

where $1 \le n_j \le n$ is the $j^{th}$ sample size, $m_0 = 0$, $m_j = n_1 + \ldots + n_j$, $1 \le j \le p$, $m_p = n$. Then, upon taking $G(y) \equiv y$, $d_{ij} \equiv x_{ij}$ in (7*), we obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \sum_{k=1}^{n_j} \{|Y_{ij} + Y_{kj} - 2t_j| - |Y_{ij} - Y_{kj}|\}, \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $Y_{ij}$ = the $i^{th}$ observation from the $j^{th}$ treatment, $1 \le j \le p$. Consequently, $\boldsymbol{\beta}^+ = (\beta_1^+, \ldots, \beta_p^+)'$, where $\beta_j^+ = \text{med}\,\{(Y_{ij} + Y_{kj})2^{-1}, 1 \le i \le k \le n_j\}$, $1 \le j \le p$. That is, in a completely randomized design with p treatments, $\boldsymbol{\beta}^+$ corresponding to the weights $\mathbf{d}_i = \mathbf{x}_i$ and $G(y) \equiv y$ is the vector of Hodges–Lehmann estimators. Similar remark applies to the randomized block, factorial and other similar designs.

The class of estimators $\boldsymbol{\beta}^+$ also includes the well celebrated *least absolute deviation* (l.a.d.) estimator. To see this, assume that *the errors are continuous.* Choose $G = \delta_0$ – the measure degenerate at 0 – in $K^+$, to obtain

$$\text{(11)} \qquad \begin{aligned} K^+(\mathbf{t}) &= \sum_{j=1}^{p} \Big[\sum_{i=1}^{n} d_{ij}\, \{I(Y_i - \mathbf{x}_i'\mathbf{t} \le 0) - I(Y_i - \mathbf{x}_i'\mathbf{t} > 0)\}\Big]^2 \\ &= \sum_{j=1}^{p} \Big(\sum_{i=1}^{n} d_{ij}\, \text{sgn}(Y_i - \mathbf{x}_i'\mathbf{t})\Big)^2, \text{ w.p.1}, \qquad \forall\ \mathbf{t} \in \mathbb{R}^p. \end{aligned}$$

Upon choosing $\mathbf{d}_i \equiv \mathbf{x}_i$, one sees that the r.h.s. of (11) is precisely the square of the norm of a.e. differential of the sum of absolute deviations $\mathscr{D}(\mathbf{t}) := \Sigma_i\, |Y_i - \mathbf{x}_i'\mathbf{t}|$, $\mathbf{t} \in \mathbb{R}^p$. Clearly the minimizer of $\mathscr{D}(\mathbf{t})$ is also a minimizer of $K^+(\mathbf{t})$ of (11).

Any one of the expressions among (7), (7′) or (7*) may be used to compute $\boldsymbol{\beta}^+$ for a general G. From these expressions it becomes apparent that the computation of $\boldsymbol{\beta}^+$ is similar to the computation of maximum likelihood estimators. It is also apparent from the above discussion that both classes $\{\hat{\boldsymbol{\beta}}\}$ and $\{\boldsymbol{\beta}^+\}$ include rather interesting estimators. On the one hand we have a smooth unbounded G, v.i.z., $G(y) \equiv y$, giving rise to Hodges–Lehmann type estimators and on the other hand a highly discrete G, v.i.z., $G = \delta_0$, giving rise to the l.a.d.e.. Any large sample theory should be general enough to cover both of these cases.

We now address the question of the *existence* of these estimators in the case $p = 1$. As before when $p = 1$, we write unbold letters for scalars and $d_i$, $x_i$ for $d_{i1}$, $x_{i1}$, $1 \le i \le n$. Before stating the result we need to define

$$\Gamma(y) := \Sigma_i\, I(x_i = 0)\, d_i\{I(Y_i \le y) - I(-Y_i < y)\}, \qquad y \in \mathbb{R}.$$

Arguing as for (7) we obtain, with $b = \max\{Y_i, -Y_i;\ 1 \le i \le n\}$,

$$\text{(12)} \qquad \int |\Gamma|\, dG \le \Sigma_i\, I(x_i = 0)|d_i|[G(b_-)\text{-}G(Y_{i-}) + G(b_-)\text{-}G(-Y_i)] < \infty.$$

Moreover, directly from (7) we can conclude that

$$\text{(13)} \qquad \int \Gamma^2\, dG < \infty.$$

Both (12) and (13) hold for all $n \ge 1$, for every sample $\{Y_i\}$ and for all real numbers $\{d_i\}$.

**Lemma 5.3.1.** *Assume that* (1.1.1) *with* $p = 1$ *holds. In addition, assume that either*

$$\text{(14a)} \qquad d_i x_i \ge 0, \quad \forall\ 1 \le i \le n, \quad \textit{or} \quad \text{(14b)} \qquad d_i x_i \le 0, \quad \forall\ 1 \le i \le n.$$

*Then a minimizer of* $K^+$ *exists if either Case* 1: $G(\mathbb{R}) = \infty$, *or Case* 2: $G(\mathbb{R}) < \infty$ *and* $d_i = 0$ *whenever* $x_i = 0$, $1 \le i \le n$.

*If* G *is continuous then a minimizer is measurable.*

**Proof.** The proof uses Fatou's Lemma and the D.C.T. Specialize (5.2.19) to the case $p = 1$ to obtain

$$K^+(t) = \int [\Sigma_i\, d_i\{I(Y_i \le y + x_i t) - I(-Y_i < y - x_i t)\}]^2\, dG(y).$$

Let $\mathcal{K}^+(y, t)$ denote the integrand without the square. Then

$$\mathcal{K}^+(y, t) = \Gamma(y) + \mathcal{K}^*(y, t),$$

where

$$\text{(15)} \quad \mathcal{K}^*(y, t) = \Sigma_i\, I(x_i > 0)\, d_i\{I(Y_i \le y + x_i t) - I(-Y_i < y - x_i t)\} + \\ + \Sigma_i\, I(x_i < 0)\, d_i\{I(Y_i \le y + x_i t) - I(-Y_i < y - x_i t)\}.$$

Clearly, $\forall\ y, t \in \mathbb{R}$,

$$|\mathcal{K}^*(y, t)| \le \Sigma_i\, I(x_i \ne 0)|d_i| =: \alpha, \qquad \text{say.}$$

Hence

$$(16) \qquad \Gamma(y) - \alpha \le \mathcal{K}^{+}(y, t) \le \Gamma(y) + \alpha, \qquad \forall\ y, t \in \mathbb{R}.$$

Suppose that **(14a) holds**. Then, from (15) it follows that $\forall\ y \in \mathbb{R}$,

$$\mathcal{K}^{*}(y, t) \longrightarrow \pm\, \alpha \quad \text{as} \quad t \longrightarrow \pm\, \infty,$$

so that $\forall\ y \in \mathbb{R}$,

$$(17) \qquad \mathcal{K}^{+}(y, t) \longrightarrow \Gamma(y) \pm \alpha, \ \text{as}\ t \longrightarrow \pm\, \infty.$$

Now consider **Case 1**. If $\alpha = 0$ then either all $x_i \equiv 0$ or $d_i = 0$ for those $i$ for which $x_i \neq 0$. In either case one obtains from (13) and (16) that $\forall\ t \in \mathcal{R}$, $K^{+}(t) = \int \Gamma^2\, dG < \infty$, and hence a minimizer trivially exists.

If $\alpha > 0$ then, from (12) and (13) it follows that $\int (\Gamma(y) \pm \alpha)^2\, dG(y) = \infty$, and by (16) and the Fatou Lemma, $\liminf_{t \longrightarrow \pm\infty} K^{+}(t) = \infty$. On the other hand by (7), $K^{+}(t)$ is a finite number for every real $t$, and hence a minimizer exists.

Next, consider **Case 2**. Here, clearly $\Gamma \equiv 0$. From (16), we obtain

$$\{\mathcal{K}^{+}(y, t)\}^2 \le \alpha^2, \qquad \forall\ y, t \in \mathbb{R},$$

and hence

$$K^{+}(t) \le \alpha^2 G(\mathbb{R}), \qquad \forall\ t \in \mathbb{R}.$$

By (17), $\mathcal{K}^{+}(y, t) \longrightarrow \pm\, \alpha$, as $t \longrightarrow \pm\, \infty$. By the D.C.T. we obtain

$$K^{+}(t) \longrightarrow \alpha^2 G(\mathbb{R}), \quad \text{as} \quad |t| \longrightarrow \infty,$$

thereby proving the existence of a minimizer of $K^{+}$ in Case 2.

The continuity of $G$ together with (7*) shows that $K^{+}$ is a continuous function on $\mathbb{R}$ thereby ensuring the measurability of a minimizer, by Corollary 2.1 of Brown and Purves (1973). This completes the proof in the case of (14a). It is exactly similar when (14b) holds, hence no details will be given for that case. □

**Remark 5.3.1.** Observe that in some cases minimizers of $K^{+}$ could be measurable even if $G$ is not continuous. For example, in the case of l.a.d. estimator, $G$ is degenerate at $0$ yet a measurable minimizer exists.

The above proof is essentially due to Dhar (1991a). Dhar (1991b) gives proofs of the existence of classes of estimators $\{\hat{\beta}\}$ and $\{\beta^+\}$ of (5.2.11) and (5.2.20) for $p \geq 1$, among other results. These proofs are somewhat complicated and will not be reproduced here. In both of these papers Dhar carries out some finite sample simulation studies and concludes that both, $\hat{\beta}$ and $\beta^+$ corresponding to $G(y) \equiv y$, show some superiority over some of the well known estimators.

Note that (14a) is *a priori* satisfied by the weights $d_i \equiv x_i$. □

Now we discuss $\bar{\beta}$ of (5.2.14). We rewrite

$$Q(\mathbf{t}) = n^{-1}\sum_i \sum_j L_{ij} \int \{I(Y_i \leq y) - H_i(y - \mathbf{x}_i'\mathbf{t})\}\{I(Y_j \leq y) - H_j(y - \mathbf{x}_j'\mathbf{t})\} dG(y)$$

where $L_{ij} = 1 - L((i \vee j)n^{-1})$, $1 \leq i, j \leq n$. Differentiating $Q$ w.r.t. $\mathbf{t}$ under the integral sign (which can be easily justified assuming $H_i$ has density $h_i$ and some other mild conditions) we obtain

$$\text{(18)} \quad \dot{Q}(\mathbf{t}) = 2n^{-1} \Sigma_i \Sigma_j L_{ij} \int \{I(Y_i \leq y) - H_i(y - \mathbf{x}_i'\mathbf{t})\} h_i(y - \mathbf{x}_j'\mathbf{t})\, dG(y)\, \mathbf{x}_j.$$

Specialize this to the case $G(y) \equiv y$, $L(s) \equiv s$, $p = 1$, $x_i \equiv 1$ and integrate by parts, to obtain

$$\begin{aligned}\dot{Q}(t) &= -2n^{-2} \Sigma_i \Sigma_j \min(n - i, n - j)\{H_i(Y_i - t) - 1/2\} \\ &= -n^{-2} \Sigma_i (n - i)(n + i - 1) \{H_i(Y_i - t) - 1/2\}.\end{aligned}$$

Now suppose further that $H_i \equiv F$. Then $\bar{\beta}$ is a solution $t$ of

$$\text{(19)} \qquad \Sigma_i (n - i)(n + i - 1)\{F(Y_i - t) - 1/2\} = 0.$$

Compare this $\bar{\beta}$ with $\hat{\theta}$ of (2). Clearly $\bar{\beta}$ given by (19) is a weighted M-estimator of the location parameter whereas $\hat{\theta}$ given by (7) is an ordinary M-estimator. Of course, if in (18) we choose $L(s) = I(s \geq 1)$, $p = 1$, $x_i \equiv 1$, $G(y) = y$ then $\hat{\theta} = \bar{\beta}$. In general $\bar{\beta}$ may be obtained as a solution of $\dot{Q}(\mathbf{t}) = \mathbf{0}$.

Next, consider $\beta^*$ of (5.2.23). For the time being focus on the case $p = 1$ and $d_i \equiv x_i - \bar{x}$. Assume, without loss of generality, that the data is so arranged that $x_1 \leq x_2 \leq \ldots \leq x_n$. Let $\mathscr{S} := \{(Y_j - Y_i)/(x_j - x_i);\ i < j,\ x_i < x_j\}$, $t_0 := \min\{t;\ t \in \mathscr{S}\}$ and $t_1 := \max\{t;\ t \in \mathscr{S}\}$. Then for $x_i < x_j$, $t < t_0$ implies $t < (Y_j - Y_i)/(x_j - x_i)$ so that $R_{it} < R_{jt}$. In other words the residuals $\{Y_j - tx_j;\ 1 \leq j \leq n\}$ are naturally ordered for all $t < t_0$, w.p.1., assuming the continuity of the errors. Hence, with $T(s, t)$ denoting the $T_{1d}(s, t)$ of (22), we obtain for $t < t_0$,

$$T(s, t) = \sum_{i=1}^{k} d_i, \qquad k/n \le s < (k+1)/n, \quad 1 \le k \le n-1,$$
$$= 0, \qquad 0 \le s < 1/n, \quad s = 1.$$

Hence,

$$K^*(t) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2, \qquad t < t_0.$$

where $\omega_k = [L((k+1)/n) - L(k/n)]$, $1 \le k \le n-1$. Consequently

$$K^*(t_0-) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2.$$

Similarly using the fact $\Sigma_i d_i = 0$, one obtains

$$K^*(t) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 = K^*(t_{1+}), \quad t > t_1.$$

As t crosses over $t_0$ only one pair of adjacent residuals change their ranks. Let $x_j < x_{j+1}$ denote their respective regression constants. Then

$$\begin{aligned} K^*(t_0-) - K^*(t_0+) &= \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 - \\ &\quad - [\sum_{\substack{k=1 \\ k\neq j}}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 + \omega_j \{d_{j+1} + \sum_{i=1}^{j-1} d_i\}^2] \\ &= \omega_k [\{ \sum_{i=1}^{j} d_i\}^2 - \{d_{j+1} + \sum_{i=1}^{j-1} d_i\}^2]. \end{aligned}$$

But $x_1 \le x_2 \le \ldots \le x_n$, $x_j < x_{j+1}$ and $\Sigma_i d_i = 0$ imply

$$\sum_{i=1}^{j} d_i < d_{k+1} + \sum_{i=1}^{j-1} d_i \le 0.$$

Hence $K^*(t_0-) > K^*(t_0+)$. Similarly it follows that $K^*(t_{1+}) > K^*(t_1-)$. Consequently, $\beta_1$ and $\beta_2$ are finite, where

$$\beta_1 := \min\{t \in \mathscr{S}, K^*(t_+) = \inf_{\Delta \in \mathscr{S}^c} K^*(\Delta)\},$$
$$\beta_2 := \max\{t \in \mathscr{S}, K^*(t_-) = \inf_{\Delta \in \mathscr{S}^c} K^*(\Delta)\},$$

and where $\mathscr{S}^c$ denotes the complement of $\mathscr{S}$. Then $\beta^*$ can be uniquely defined by the relation $\beta^* = (\beta_1 + \beta_2)/2$.

This $\beta^*$ corresponding to $L(s) \equiv s$ was studied by Williamson (1979, 1982). In general this estimator is asymptotically relatively more efficient than Wilcoxon type R-estimators as will be seen later on in Section 5.6.

There does not seem to be such a nice characterization for $p \geq 1$ and general $\mathbf{D}$ satisfying (5.2.21). However, proceeding as in the derivation of (6), a computational formula for $K^*$ of (5.2.22) can be obtained to be

$$K^*(\mathbf{t}) = -2 \sum_{k=1}^{p} \Sigma_i \Sigma_j\, d_{ij} d_{jk}\, |L((R_{it}/n)-) - L((R_{jt}/n)-)|. \tag{20}$$

This formula is valid for a general $\sigma$–finite measure $L$ and can be used to compute $\boldsymbol{\beta}^*$.

We now turn to the m.d. estimator defined at (5.2.26) and (5.2.27). Let $d_i \equiv x_i - \bar{x}$. The first observation one makes is that for $t \in \mathbb{R}$,

$$\mathcal{D}_n(t) := \sup_{y \in \mathbb{R}} |\sum_{i=1}^{n} d_i I(Y_i \leq y + td_i)| = \sup_{0 \leq s \leq 1} |\sum_{i=1}^{n} d_i I(R_{it} \leq ns)|.$$

Proceeding as in the above discussion pertaining to $\beta^*$, assume, without loss of generality, that the data is so arranged that $x_1 \leq x_2 \leq \ldots \leq x_n$ so that $d_1 \leq d_2 \leq \ldots \leq d_n$. Let $\mathcal{A}_1 := \{(Y_j - Y_i)/(d_j - d_i);\ d_i < 0,\ d_j \geq 0,\ 1 \leq i < j \leq n\}$. It can be proved that $\mathcal{D}_n^+$ ($\mathcal{D}_n^-$) is a left continuous non–decreasing (right continuous non–increasing) step function on $\mathbb{R}$ whose points of discontinuity are a subset of $\mathcal{A}_1$. Moreover, if $-\infty = t_0 < t_1 \leq t_2 \leq \ldots \leq t_m < t_{m+1} = \infty$ denote the ordered members of $\mathcal{A}_1$, then $\mathcal{D}_n^+(t_1-) = 0 = \mathcal{D}_n^-(t_{m}+)$ and $\mathcal{D}_n^+(t_{m}+) = \Sigma_i\, d_i^+ = \mathcal{D}_n^-(t_1-)$, where $d_i^+ \equiv \max(d_i, 0)$. Consequently, the following entities are finite:

$$\beta_{s1} := \inf\{t \in \mathbb{R};\ \mathcal{D}_n^+(t) \geq \mathcal{D}_n^-(t)\}, \qquad \beta_{s2} := \sup\{t \in \mathbb{R};\ \mathcal{D}_n^+(t) \leq \mathcal{D}_n^-(t)\}.$$

Note that $\beta_{s2} \geq \beta_{s1}$ w.p.1.. One can now take $\beta_s = (\beta_{s1} + \beta_{s2})/2$.

Williamson (1979) provides the proofs of the above claims and obtains the asymptotic distribution of $\beta_s$. This estimator is the precise generalization of the m.d. estimator of the two sample location parameter of Rao, Schuster and Littell (1975). Its asymptotic distribution is the same as that of their estimator.

We shall now discuss some distributional properties of the above m.d. estimators. To facilitate this discussion let $\tilde{\boldsymbol{\beta}}$ denote any one of the estimators defined at (5.2.11), (5.2.20), (5.2.23) and (5.2.27). As in Section 4.3, we shall write $\tilde{\beta}(\mathbf{X}, \mathbf{Y})$ to emphasize the dependence on the data $\{(\mathbf{x}_i',\ Y_i);\ 1 \leq i \leq n\}$. It also helps to think of the defining distances $K$, $K^+$, etc. as functions of residuals. Thus we shall some times write $K(\mathbf{Y} - \mathbf{X}\mathbf{t})$ etc. for

$K(\mathbf{t})$ etc. Let $\tilde{K}$ stand for either $K$ or $K^+$ or $K^*$ of (5.2.10), (5.2.19) and (5.2.22). To begin with, observe that

(21) $$\tilde{K}(\mathbf{t}-\mathbf{b}) = \tilde{K}(\mathbf{Y}+\mathbf{X}\mathbf{b}-\mathbf{X}\mathbf{t}), \qquad \forall\ \mathbf{t}, \mathbf{b} \in \mathbb{R}^p,$$

so that

(22) $$\tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}+\mathbf{X}\mathbf{b}) = \tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) + \mathbf{b}, \qquad \forall\ \mathbf{b} \in \mathbb{R}^p.$$

Consequently, the distribution of $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}$ does not depend on $\boldsymbol{\beta}$.

The distance measure $Q$ of (5.2.13) does not satisfy (21) and hence the distribution of $\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}$ will generally depend on $\boldsymbol{\beta}$.

In general, the classes of estimators $\{\hat{\boldsymbol{\beta}}\}$ and $\{\boldsymbol{\beta}^+\}$ are not scale invariant. However, as can be readily seen from (6) and (7), the class $\{\hat{\boldsymbol{\beta}}\}$ corresponding to $G(y) \equiv y$, $H_i \equiv F$ and those $\{D\}$ that satisfy (5.2.21) and the class $\{\boldsymbol{\beta}^+\}$ corresponding to $G(y) \equiv y$ and general $\{D\}$ are scale invariant in the sense of (4.3.2).

An interesting property of all of the above m.d. estimators is that they are invariant under nonsingular transformation of the design matrix $\mathbf{X}$. That is,

$$\tilde{\boldsymbol{\beta}}(\mathbf{X}\mathbf{B}, \mathbf{Y}) = \mathbf{B}^{-1}\tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) \text{ for every } p\times p \text{ nonsingular matrix } \mathbf{B}.$$

A similar statement holds for $\bar{\boldsymbol{\beta}}$.

We shall end this section by discussing the *symmetry* property of these estimators. In the following lemma it is implicitly assumed that all integrals involved are finite. Some sufficient conditions for that to happen will unfold as we proceed in this chapter.

**Lemma 5.3.2.** *Let* (1.1.1) *hold with the actual and the modeled d.f. of* $e_i$ *equal to* $H_i$, $1 \le i \le n$.

(i) *If either*

(ia) $\{H_i,\ 1 \le i \le n\}$ *and* $G$ *are symmetric around* 0 *and* $\{H_i,\ 1 \le i \le n\}$ *are continuous,*

*or*

(ib) $d_{ij} = -\, d_{n-i+1,j}$, $x_{ij} = -x_{n-i+1,j}$ *and* $H_i \equiv F$ $\forall$ $1 \le i \le n$, $1 \le j \le p$,

*then*

$\hat{\boldsymbol{\beta}}$ *and* $\boldsymbol{\beta}^*$ *are symmetrically distributed around* $\boldsymbol{\beta}$, *whenever they exist uniquely.*

(ii) *If* $\{H_i,\ 1 \le i \le n\}$ *and* G *are symmetric around* 0 *and either* $\{H_i,\ 1 \le i \le n\}$ *are continuous or* G *is continuous,*

*then*

$\boldsymbol{\beta}^+$ *is symmetrically distributed around* $\boldsymbol{\beta}$*, whenever it exists uniquely.*

**Proof.** In view of (22) there is no loss of generality in assuming that the true $\boldsymbol{\beta}$ is $\mathbf{0}$.

Suppose that **(ia) holds.** Then $\hat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) =_d \hat{\boldsymbol{\beta}}(\mathbf{X}, -\mathbf{Y})$. But, by definition (5.2.11), $\hat{\boldsymbol{\beta}}(\mathbf{X}, -\mathbf{Y})$ is the minimizer of $K(-\mathbf{Y} - \mathbf{X}\mathbf{t})$ w.r.t. $\mathbf{t}$. Observe that $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$K(-\mathbf{Y} - \mathbf{X}\mathbf{t}) = \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(-Y_i \le y + \mathbf{x}_i'\mathbf{t}) - H_i(y)\}]^2\, dG(y)$$

$$= \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{1 - I(Y_i < -y - \mathbf{x}_i'\mathbf{t}) - H_i(y)\}]^2\, dG(y)$$

$$= \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(Y_i < y - \mathbf{x}_i'\mathbf{t}) - H_i(y-)\}]^2\, dG(y)$$

by the symmetry of $\{H_i\}$ and G. Now use the continuity of $\{H_i\}$ to conclude that, w.p.1.,

$$K(-\mathbf{Y} - \mathbf{X}\mathbf{t}) = K(\mathbf{Y} + \mathbf{X}\mathbf{t}), \qquad \forall\, \mathbf{t} \in \mathbb{R}^p,$$

so that $\hat{\boldsymbol{\beta}}(\mathbf{X}, -\mathbf{Y}) = -\hat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y})$, w.p.1, and the claim follows because $-\hat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) = \operatorname{argmin}\{K(\mathbf{Y} + \mathbf{X}\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}$.

Now suppose that **(ib) holds.** Then

$$\begin{aligned}
&K(\mathbf{Y} + \mathbf{X}\mathbf{t}) \\
&= \sum_{j=1}^{p} \int [\Sigma_i\, d_{n-i+1,j}\{I(Y_i \le y + \mathbf{x}_{n-i+1}'\mathbf{t}) - F(y)\}]^2\, dG(y) \\
&=_d \sum_{j=1}^{p} \int [\Sigma_i\, d_{n-i+1,j}\{I(Y_{n-i+1} \le y + \mathbf{x}_{n-i+1}'\mathbf{t}) - F(y)\}]^2\, dG(y) \\
&= K(\mathbf{Y} - \mathbf{X}\mathbf{t}), \qquad \forall\ \mathbf{t} \in \mathbb{R}^p.
\end{aligned}$$

This shows that $-\hat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) =_d \hat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y})$ as required. The proof for $\boldsymbol{\beta}^*$ is similar.

**Proof of (ii).** Again, $\boldsymbol{\beta}^{+}(\mathbf{X}, \mathbf{Y}) \underset{d}{=} \boldsymbol{\beta}^{+}(\mathbf{X}, -\mathbf{Y})$, because of the symmetry of $\{H_i\}$. But,

$$
\begin{aligned}
&K^{+}(-\mathbf{Y} - \mathbf{X}\mathbf{t}) \\
&= \sum_{j=1}^{p} \int [\Sigma_i \, d_{ij} I(-Y_i \leq y + \mathbf{x}_i'\mathbf{t}) - 1 + I(-Y_i \leq -y + \mathbf{x}_i'\mathbf{t})\}]^2 \, dG(y) \\
&= \sum_{j=1}^{p} \int [\Sigma_i \, d_{ij} \, \{I(Y_i < y + \mathbf{x}_i'\mathbf{t}) - 1 + I(Y_i < -y + \mathbf{x}_i'\mathbf{t})\}]^2 \, dG(y) \\
&= K^{+}(\mathbf{Y} + \mathbf{X}\mathbf{t}), \qquad\qquad \forall \ \mathbf{t} \in \mathbb{R}^p,
\end{aligned}
$$

w.p.1, if either $\{H_i\}$ or $G$ are continuous. □

## 5.4. ASYMPTOTICS OF MINIMUM DISPERSION ESTIMATORS: A GENERAL CASE

This section gives a general overview of an asymptotic theory useful in inference based on minimizing an objective function of the data and parameter in general models. It is a self contained section of broad interest.

In an inferential problem consisting of a vector of n observations $\zeta_n = (\zeta_{n1}, \ldots, \zeta_{nn})'$, not necessarily independent, and a p-dimensional parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, an estimator of $\boldsymbol{\theta}$ is often based on an objective function $M_n(\zeta_n, \boldsymbol{\theta})$, herein called *dispersion.* In this section an estimator of $\boldsymbol{\theta}$ obtained by minimizing $M_n(\zeta_n, \cdot)$ will be called *minimum dispersion estimator.*

Typically the sequence of dispersion $M_n$ admits the following approximate quadratic structure. Writing $M_n(\boldsymbol{\theta})$ for $M_n(\zeta_n, \boldsymbol{\theta})$, often it turns out that $M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)$, under $\boldsymbol{\theta}_0$, is asymptotically like a quadratic form in $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, for $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}_0$ in a certain sense, with the coefficient of the linear term equal to a random vector which is asymptotically normally distributed. This approximation in turn is used to obtain the asymptotic distribution of minimum dispersion estimators.

The two classical examples of the above type are Gauss's least square and Fisher's maximum likelihood estimators. In the former the dispersion $M_n$ is the error sum of squares while in the latter $M_n$ equals $-\log L_n$, $L_n$ denoting the likelihood function of $\boldsymbol{\theta}$ based on $\zeta_n$. In the least squares method, $M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)$ is exactly quadratic in $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, uniformly in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. The random vector appearing in the linear term is typically asymptotically normally distributed. In the likelihood method, the well celebrated locally asymptotically normal (l.a.n.) models of Le Cam (1960, 1986) obey the above type of approximate quadratic structure. Other well known examples include the least absolute deviation and the minimum chi-square estimators.

The main purpose of this section is to unify the basic structure of asymptotics underlying the minimum dispersion estimators by exploiting the above type of common asymptotic quadratic structure inherent in most of the dispersions.

We now formulate general conditions for a given dispersion to be uniformly locally asymptotically quadratic (u.l.a.q.d.). Accordingly, let $\Omega$ be an open subset of $\mathbb{R}^p$ and $M_n$, $n \geq 1$, be a sequence of real valued functions defined on $\mathbb{R}^n \times \Omega$ such that $M_n(\cdot, \boldsymbol{\theta})$ is measurable for each $\boldsymbol{\theta}$. We shall often suppress the $\zeta_n$ coordinate in $M_n$ and write $M_n(\boldsymbol{\theta})$ for $M_n(\zeta_n, \boldsymbol{\theta})$.

In order to state general conditions we need to define a sequence of neighborhoods $N_n(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} \in \Omega, |\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \leq B\}$, where $\boldsymbol{\theta}_0$ is a fixed parameter value in $\Omega$, B is a finite number and $\{\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)\}$ is a sequence of p×p symmetric positive definite matrices with norms $\|\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)\|$ tending to infinity. Since $\boldsymbol{\theta}_0$ is fixed, write $\boldsymbol{\delta}_n$, $N_n$ for $\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)$, $N_n(\boldsymbol{\theta}_0)$, respectively. Similarly, let $P_n$ denote the probability distribution of $\zeta_n$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

**Definition 5.4.1.** A sequence of dispersions $\{M_n(\boldsymbol{\theta}), \boldsymbol{\theta} \in N_n\}$, $n \geq 1$, is said to be u.l.a.q. (*uniformly locally asymptotically quadratic*) if it satisfies condition $(A\tilde{1}) - (A\tilde{3})$ given below.

$(A\tilde{1})$ There exist a sequence of p×1 random vector $S_n(\boldsymbol{\theta}_0)$ and a sequence of p×p, possibly random, matrices $W_n(\boldsymbol{\theta}_0)$, such that, for every $0 < B < \infty$, and for all $\boldsymbol{\theta} \in N_n$,

$$M_n(\boldsymbol{\theta}) = M_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' S_n(\boldsymbol{\theta}_0) + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' W_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \bar{o}_p(1),$$

where " $\bar{o}_p(1)$" is a sequence of stochastic processes in $\boldsymbol{\theta}$ converging to zero, *uniformly in* $\boldsymbol{\theta} \in N_n$, in $P_n$-probability.

$(A\tilde{2})$ There exists a p×p non-singular, possibly random, matrix $\mathbf{W}(\boldsymbol{\theta}_0)$ such that

$$\boldsymbol{\delta}_n^{-1} W_n(\boldsymbol{\theta}_0) \boldsymbol{\delta}_n^{-1} = \mathbf{W}(\boldsymbol{\theta}_0) + o_p(1), \qquad (P_n).$$

$(A\tilde{3})$ There exists a p×1 r.v. $Y(\boldsymbol{\theta}_0)$ such that

$$\mathscr{L}_n\left(\boldsymbol{\delta}_n^{-1} S_n(\boldsymbol{\theta}_0), \boldsymbol{\delta}_n^{-1} W_n(\boldsymbol{\theta}_0) \boldsymbol{\delta}_n^{-1}\right) \Rightarrow \mathscr{L}\left(Y(\boldsymbol{\theta}_0), \mathbf{W}(\boldsymbol{\theta}_0)\right)$$

where $\mathscr{L}_n$, $\mathscr{L}$ denote joint probability distributions under $P_n$ and in the limit, respectively.

Denote the conditions $(A\tilde{1})$, $(A\tilde{2})$ by (A1) and (A2), respectively, whenever $\mathbf{W}$ is *non-random* in these conditions. A sequence of dispersions $\{M_n\}$ is called *uniformly locally asymptotically **normal** quadratic* (u.l.a.n.q.) if (A1), (A2) hold and if (A3), instead of $(A\tilde{3})$, holds, where (A3) is as follows:

(A3) There exists a positive definite $p \times p$ matrix $\Sigma(\theta_0)$ such that

$$\delta_n^{-1} S_n(\theta_0) \underset{d}{\longrightarrow} N(0, \Sigma(\theta_0)), \qquad (P_n).$$

If $(\tilde{A1})$ holds *without* the *uniformity* requirement and $(\tilde{A2})$, $(\tilde{A3})$ hold then we call the given sequence $M_n$ *locally asymptotically quadratic* (l.a.q.). If (A1) holds *without* the *uniformity* requirement and (A2), (A3) hold then the given sequence $M_n$ is called *locally asymptotically normal quadratic.*

In the case $M_n(\theta) = -\ell n\ L_n(\theta)$, the conditions non–uniform (A1), (A2), (A3) with $\|\delta_n\| = O(n^{1/2})$, determine the well celebrated l.a.n. models of Le Cam (1960, 1986). For this particular case, $W(\theta_0)$, $\Sigma(\theta_0)$ and the limiting Fisher information matrix $F(\theta_0)$, whenever it exists, are the same.

In the above general formulation, $M_n$ is an arbitrary dispersion satisfying $(\tilde{A1})$ – $(\tilde{A3})$ or (A1) – (A3). In the latter the three matrices $W(\theta_0)$, $\Sigma(\theta_0)$ and $F(\theta_0)$ are not necessarily identical. The l.a.n.q. dispersions can thus be viewed as a generalization of the l.a.n. models.

Typically in the classical i.i.d. setup the normalizing matrix $\delta_n$ is of the order square root of n whereas in the linear regression model (1.1.1) it is of the order $(X'X)^{1/2}$. In general $\delta_n$ will depend on $\theta_0$ and is determined by the order of the asymptotic magnitude of $S_n(\theta_0)$.

An example where the full strength of $(\tilde{A1})$ – $(\tilde{A3})$ is realized is obtained by considering the least square dispersion in an explosive autoregression model where for some $|\rho| > 1$, $X_i = \rho X_{i-1} + e_i$, $i \geq 1$, and where $\{e_i, i \geq 1\}$ are i.i.d. r.v.'s. For details see Koul and Pflug (1990).

We now turn to the asymptotic distribution of the minimum dispersion estimators. Let $\{M_n\}$ be a sequence of u.l.a.q.d.'s. Define

$$\hat{\theta}_n = \operatorname{argmin}\{M_n(t),\ t \in \Omega\}. \tag{1}$$

Our goal is to investigate the asymptotic behavior of $\hat{\theta}_n$ and $M_n(\hat{\theta}_n)$. Akin to the study of the asymptotic distribution of m.l.e.'s, we must first ensure that there is a $\hat{\theta}_n$ satisfying (1) such that

$$|\delta_n(\hat{\theta}_n - \theta_0)| = O_p(1). \tag{2}$$

Unfortunately the u.l.a.q. assumptions are not enough to guarantee (2). One set of additional assumptions that ensures (2) is the following.

(A4) $\forall\ \epsilon > 0\ \exists$ a $0 < Z_\epsilon < \infty$ and $N_{1\epsilon}$ such that

$$P_n(|M_n(\theta_0)| \leq Z_\epsilon) \geq 1 - \epsilon, \qquad \forall\ n \geq N_{1\epsilon}.$$

(A5) $\forall\ \epsilon > 0$ and $0 < \alpha < \infty$, $\exists$ an $N_{2\epsilon}$ and a b (depending on $\epsilon$ and $\alpha$) such that

$$P_n\Big( \inf_{\|\delta_n(t-\theta_0)\|>b} M_n(t) \geq \alpha\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N_{2\epsilon}.$$

It is convenient to let

$$Q_n(\theta, \theta_0) := (\theta-\theta_0)'S_n(\theta_0) + (1/2)(\theta-\theta_0)'W_n(\theta_0)(\theta-\theta_0), \qquad \theta \in \mathbb{R}^p,$$

and $\tilde{\theta}_n := \operatorname{argmin}\{Q_n(\theta, \theta_0),\ \theta \in \mathbb{R}^p\}$. Clearly, $\tilde{\theta}_n$ must satisfy the relation

$$\mathscr{B}_n\, \delta_n(\tilde{\theta}_n - \theta_0) = -\delta_n^{-1}S_n(\theta_0). \tag{3}$$

where $\mathscr{B}_n := \delta_n^{-1}W_n\delta_n^{-1}$, where $W_n = W_n(\theta_0)$.

Some generality is achieved by making the following assumption.

$$\|\delta_n(\tilde{\theta}_n - \theta_0)\| = O_p(1). \tag{A6}$$

Note that $(\mathrm{A}\tilde{2})$ and $(\mathrm{A}\tilde{3})$ imply (A6). We now state and prove

**Theorem 5.4.1.** *Let the dispersions* $M_n$ *satisfy* $(\mathrm{A}\tilde{1})$, (A4) – (A6). *Then, under* $P_n$,

$$|(\hat{\theta}_n - \tilde{\theta}_n)'\delta_n\mathscr{B}_n\delta_n(\hat{\theta}_n - \tilde{\theta}_n)| = o_p(1), \tag{4}$$

$$\inf_{\theta\in\Omega} M_n(\theta) - M_n(\theta_0) = -(1/2)(\tilde{\theta}_n - \theta_0)'W_n(\tilde{\theta}_n - \theta_0) + o_p(1). \tag{5}$$

*Consequently, if* (A6) *is replaced by* $(\mathrm{A}\tilde{2})$ *and* $(\mathrm{A}\tilde{3})$, *then*

$$\delta_n(\hat{\theta}_n - \theta_0) \xrightarrow[d]{} \{W(\theta_0)\}^{-1}Y(\theta_0), \tag{6}$$

*and*

$$\inf_{\theta\in\Omega} M_n(\theta) - M_n(\theta_0) = -(1/2)\, S_n'(\theta_0)\delta_n^{-1}\mathscr{B}_n^{-1}\delta_n^{-1}S_n(\theta_0) + o_p(1). \tag{7}$$

*If, instead of* $(\mathrm{A}\tilde{1})$ – $(\mathrm{A}\tilde{3})$, $M_n$ *satisfies* (A1) – (A3), *and if* (A4) *and* (A5) *hold then also* (4) – (7) *hold and*

$$\delta_n(\hat{\theta}_n - \theta_0) \xrightarrow[d]{} N(0, \Gamma(\theta_0)), \tag{8}$$

*where* $\Gamma(\theta_0) = \{W(\theta_0)\}^{-1}\Sigma(\theta_0)\{W(\theta_0)\}^{-1}$.

**Proof.** Let $Z_\epsilon$ be as in (A4). Choose an $\alpha > Z_\epsilon$ in (A5). Then

$$\Big[|M_n(\theta_0)| \le Z_\epsilon, \inf_{|h|>b} M_n(\theta_0 + \delta_n^{-1}h) \ge \alpha\Big]$$

$$\subset \Big[\inf_{|h|\le b} M_n(\theta_0 + \delta_n^{-1}h) \le Z_\epsilon, \inf_{|h|>b} M_n(\theta_0 + \delta_n^{-1}h) \ge \alpha\Big]$$

$$\subset \Big[\inf_{|h|>b} M_n(\theta_0 + \delta_n^{-1}h) > \inf_{|h|\le b} M_n(\theta_0 + \delta_n^{-1}h)\Big].$$

Hence by (A4) and (A5), for any $\epsilon > 0$ there exists a b (now depending only on $\epsilon$) such that $\forall\ n \ge N_{1\epsilon} \vee N_{2\epsilon}$,

(9) $$P_n\big(\inf_{|h|>b} M_n(\theta_0 + \delta_n^{-1}h) > \inf_{|h|\le b} M_n(\theta_0 + \delta_n^{-1}h)\big) \ge 1 - \epsilon,$$

This in turn ensures the validity of (2). Having verified (2), $(\tilde{A1})$ now yields

(10) $$M_n(\hat{\theta}_n) = M_n(\theta_0) + Q_n(\hat{\theta}_n, \theta_0) + o_p(1), \qquad (P_n).$$

From (A6), the inequality

$$|\inf_{\theta\in N_n} M_n(\theta) - \inf_{\theta\in N_n} [M_n(\theta_0) + Q_n(\theta, \theta_0)]|$$

$$\le \sup_{\theta\in N_n} |M_n(\theta) - [M_n(\theta_0) + Q_n(\theta, \theta_0)]|$$

and (A1), we obtain

(11) $$M_n(\hat{\theta}_n) = M_n(\theta_0) + Q_n(\tilde{\theta}_n, \theta_0) + o_p(1), \qquad (P_n).$$

Now, (10) and (11) readily yield

$$Q_n(\hat{\theta}_n, \theta_0) = Q_n(\tilde{\theta}_n, \theta_0) + o_p(1), \qquad (P_n),$$

which is precisely equivalent to the statement (4). The calim (5) follows from (3) and (11). The rest is obvious. □

**Remark 5.4.1.** Roughly speaking, the assumption (A5) says that the smallest value of $M_n(\theta)$ outside of $N_n$ can be made asymptotically arbitrarily large with arbitrarily large probability. The assumption (A4) means that the sequence of r.v.'s $\{M_n(\theta_0)\}$ is bounded in probability. This assumption is usually verified by an application of the Markov inequality in the case $E_n|M_n(\theta_0)| = O(1)$, where $E_n$ denotes the expectation under $P_n$. In some applications $M_n(\theta_0)$ converges weakly to a r.v. which also implies (A4). Often the verification of (A5) is rendered easy by an application of a variant of the C-S inequality. Examples of this appear in the next section when dealing with m.d. estimators of the previous section. □

We now discuss the *minimum dispersion tests of simple hypothesis, briefly without many details.* Consider the simple hypothesis $H_0$: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In the special case when $M_n$ is $-\ell n\, L_n$, the likelihood ratio statistic for testing $H_0$ is given by $-2 \inf\{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0);\ \boldsymbol{\theta}\in\Omega\}$. Thus, given a general dispersion function $M_n$, we are motivated to base a test of $H_0$ on the statistic

$$(12) \qquad T_n = -2\inf\{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0);\ \boldsymbol{\theta}\in\Omega\},$$

with large values of $T_n$ being significant.

To study the asymptotic null distribution of $T_n$, note that by (7), $T_n = S_n'(\boldsymbol{\theta}_0)\boldsymbol{\delta}_n^{-1}\boldsymbol{\mathscr{B}}_n^{-1}\boldsymbol{\delta}_n^{-1}S_n(\boldsymbol{\theta}_0) + o_p(1)$, $(P_n)$. Let $Y$, $\mathbf{W}$ etc. stand for $Y(\boldsymbol{\theta}_0)$, $\mathbf{W}(\boldsymbol{\theta}_0)$, etc.

**Proposition 5.4.1.** *Under* $(A\tilde{1}) - (A\tilde{3})$, (A4), (A5), *the asymptotic null distribution of* $T_n$ *is the same as that of* $Y'\mathbf{W}^{-1}Y$.

*Under* (A1) – (A5), *the asymptotic null distribution of* $T_n$ *is the same as that of* $\mathbf{Z}'\mathbf{B}\,\mathbf{Z}$ *where* $\mathbf{Z}$ *is a* $N(0, I_{p\times p})$ *r.v. and* $\mathbf{B} = \Sigma^{1/2}\mathbf{W}^{-1}\Sigma^{1/2}$. □

**Remark 5.4.2.** Clearly if $\mathbf{W}(\boldsymbol{\theta}_0) = \Sigma(\boldsymbol{\theta}_0)$ then the asymptotic null distribution of $T_n$ is $\chi_p^2$. However, if $\mathbf{W} \neq \Sigma$, the limit distribution of $T_n$ is not a chi–square. We shall not discuss the distribution of $T_n$ under alternatives. □

A class of examples of the u.l.a.n.q.d.'s where (A1) – (A5) are satisfied with typically $\mathbf{W} \neq \Sigma$ is given by *Huber's M–dispersions* for the model (1.1.1), v.i.z.,

$$M_n(\mathbf{t}) = \Sigma_i\, \rho(Y_i - \mathbf{x}_i'\mathbf{t}), \qquad \mathbf{t}\in\mathbb{R}^p,$$

where $\rho$ is a convex function on $\mathbb{R}$ with its almost everywhere derivative $\psi$. As mentioned in Chapter 4 the estimators obtained by minimizing $M_n$ are studied extensively in the literature, see Huber (1981) and references there in. These estimators include the least square and the l.a.d. estimators of $\boldsymbol{\beta}$. Now, let $g_r(t) := \int[\psi(x) - \psi(x-t)]^r dF(x)$, $t\in\mathbb{R}$, $r = 1, 2$, and suppose that F and $\psi$ are such that $\int\psi dF = 0$, $0 < \int\psi^2 dF < \infty$, $g_1$ is continuously differentiable at 0 and that $g_2$ is continuous at 0. Then it can be shown, under (NX), that Huber's dispersion is l.a.n.q. with

$$\boldsymbol{\theta}_0 = \boldsymbol{\beta}, \quad \boldsymbol{\delta}_n = (\mathbf{X}'\mathbf{X})^{1/2}, \quad S_n(\boldsymbol{\beta}) = -\Sigma_i\, \mathbf{x}_i\, \psi(Y_i - \mathbf{x}_i'\boldsymbol{\beta}),$$

$$\mathbf{W}_n(\boldsymbol{\beta}) = \dot{g}(0)\mathbf{X}'\mathbf{X}, \quad \mathbf{W} = \dot{g}(0)I_{p\times p}, \quad \text{and} \quad \Sigma = \int\psi^2 dF\, I_{p\times p}.$$

This together with the convexity of $\rho$ and a result in Rockafeller (1970) yields that the above dispersion is u.l.a.n.q.d. See also Heiler and Weiler (1988) and Pollard (1991).

For $\rho(x) = |x|$ and $F$ continuous, $\psi(x) = \text{sgn}(x)$ and $g_r(t) = 2^r|F(t) - F(0)|$. The condition on $g_1$ now translates to the usual condition on $F$ in terms of the density $f$ at $0$. For $\rho(x) = x^2$, $\psi(x) \equiv 2x$, $g_1(t) \equiv 2t$, so that $g_1$ is trivially continuously differentiable with $\dot{g}_1(0) = 2$. Note that in general $\mathbf{W} \neq \Sigma$ unless $\dot{g}(0) = \int \psi^2\, dF$ which is the case when $\psi$ is related to the likelihood scores.

The next section is devoted to verifying (A1) – (A5) for various dispersion introduce in Section 4.2.

## 5.5. ASYMPTOTIC UNIFORM QUADRATICITY

In this section we shall give sufficient conditions under which $K_{\mathbf{D}}$, $K_{\mathbf{D}}^{+}$, of Section 5.2 will satisfy (5.4.A1), (5.4.A4), (5.4.A5) and $K_{\mathbf{D}}^{*}$ and $Q$ of Section 5.2 will satisfy (5.4.A1). As is seen from the previous section this will bring us a step closer to obtaining the asymptotic distributions of various m.d. estimators introduced in Section 5.2.

To begin with we shall focus on (5.4.A1) for $K_{\mathbf{D}}$, $K_{\mathbf{D}}^{+}$ and $K_{\mathbf{D}}^{*}$. Our basic objective is to study the asymptotic distribution of $\hat{\beta}_{\mathbf{D}}$ when the actual d.f.'s of $\{e_{ni},\ 1 \le i \le n\}$ are $\{F_{ni},\ 1 \le i \le n\}$ but we model them to be $\{H_{ni},\ 1 \le i \le n\}$. Similarly, we wish to study the asymptotic distribution of $\beta_{\mathbf{D}}^{+}$ when actually the errors may not be symmetric but we model them to be so. To achieve these objectives it is necessary to obtain the asymptotic results under as general a setting as possible. This of course makes the exposition that follows look somewhat complicated. The results thus obtained will enable us to study not only the asymptotic distributions of these estimators but also some of their robustness properties. With this in mind we proceed to state our assumptions.

(1) $\mathbf{X}$ satisfies (NX).

(2) With $\mathbf{d}_{(j)}$ denoting the $j^{th}$ column of $\mathbf{D}$, $\|\mathbf{d}_{(j)}\|^2 > 0$ for at least one j; $\|\mathbf{d}_{(j)}\|^2 = 1$ for all those $j$ for which $\|\mathbf{d}_{(j)}\|^2 > 0$, $1 \le j \le p$.

(3) $\{F_{ni},\ 1 \le i \le n\}$ admit densities $\{f_{ni},\ 1 \le i \le n\}$ w.r.t. $\lambda$.

(4) $\{G_n\}$ is a sequence in $\mathbb{D}\mathcal{I}(\mathbb{R})$.

(5) With $\mathbf{d}_{ni}' = (d_{ni1}, \dots, d_{nip})$, the ith row of $\mathbf{D}$, $1 \le i \le n$,

$$\int \Sigma_i \|\mathbf{d}_{ni}\|^2 F_{ni}(1-F_{ni})\, dG_n = O(1).$$

(6) With $\gamma_n := \Sigma_i \|\mathbf{d}_{ni}\|^2 f_{ni}$,

$$\limsup_n \int_{a_n}^{b_n} \int \gamma_n(y+x)\, dG_n(y)\, dx = 0$$

for any real sequences $\{a_n\}$, $\{b_n\}$, $a_n < b_n$, $b_n - a_n \longrightarrow 0$.

(7) With $d_{nij} = d_{nij}^{+} - d_{nij}^{-}$, $1 \le j \le p$; $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$, $\kappa_{ni} := \|\mathbf{c}_{ni}\|$, $1 \le i \le n$, $\forall\ \delta > 0$, $\forall\ \|\mathbf{v}\| \le B$,

$$\limsup_n \sum_{j=1}^{p} \int [\Sigma_i\, d_{nij}^{\pm}\{F_{ni}(y + \mathbf{v}'\mathbf{c}_{ni} + \delta\kappa_{ni}) - F_{ni}(y + \mathbf{v}'\mathbf{c}_{ni} - \delta\kappa_{ni})\}]^2\, dG_n(y) \le k\,\delta^2,$$

where $k$ is a constant not depending on $\mathbf{v}$ and $\delta$.

(8) With $\mathbf{R}_{nj} := \Sigma_i\, d_{nij}\, \mathbf{x}_{ni}\, f_{ni}$, $\boldsymbol{\nu}_{nj} := \mathbf{A}\mathbf{R}_{nj}$, $1 \le j \le p$,

$$\sum_{j=1}^{p} \int \|\boldsymbol{\nu}_{nj}\|^2 dG_n = O(1).$$

(9) With $\mu_{nj}^{o}(y, \mathbf{u}) := \Sigma_i\, d_{nij} F_{ni}(y + \mathbf{c}_{ni}'\mathbf{u})$, for each $\mathbf{u} \in \mathbb{R}^p$,

$$\sum_{j=1}^{p} \int [\mu_{nj}^{o}(y, \mathbf{u}) - \mu_{nj}^{o}(y, 0) - \mathbf{u}'\boldsymbol{\nu}_{nj}(y)]^2 dG_n(y) = o(1).$$

(10) With $m_{nj} := \Sigma_i\, d_{nij}[F_{ni} - H_{ni}]$, $1 \le j \le p$; $\mathbf{m}_{\mathbf{D}}' = (m_{n1}, \ldots, m_{np})$

$$\int \|\mathbf{m}_{\mathbf{D}}\|^2\, dG_n = O(1).$$

(11) With $\boldsymbol{\Gamma}_n'(y) := (\boldsymbol{\nu}_{n1}(y), \ldots, \boldsymbol{\nu}_{np}(y)) = \mathbf{D}'\boldsymbol{\Lambda}^*(y)\mathbf{X}\mathbf{A}$, where $\boldsymbol{\Lambda}^*$ is defined at (4.2.1), and with $\bar{\boldsymbol{\Gamma}}_n := \int \boldsymbol{\Gamma}_n\, g_n\, dG_n$, where $g_n \in L_r(G_n)$, $r = 1, 2$, $n \ge 1$, is such that $g_n > 0$,

$$0 < \liminf_n \int g_n^2\, dG_n \le \limsup_n \int g_n^2\, dG_n < \infty,$$

and such that there exists an $\alpha > 0$ satisfying

$$\liminf_n \inf\{\boldsymbol{\theta}'\bar{\boldsymbol{\Gamma}}_n\boldsymbol{\theta}\,;\ \boldsymbol{\theta} \in \mathbb{R}^p,\ \|\boldsymbol{\theta}\| = 1\} \ge \alpha.$$

(12) Either

(a) $\boldsymbol{\theta}'\mathbf{d}_{ni}\mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \ge 0\ \ \forall\ 1 \le i \le n$ and $\forall\ \boldsymbol{\theta} \in \mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$.

Or

(b) $\boldsymbol{\theta}'\mathbf{d}_{ni}\mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \le 0\ \ \forall\ 1 \le i \le n$ and $\forall\ \boldsymbol{\theta} \in \mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$.

In most of the subsequent applications of the results obtained in this section, the sequence of integrating measures $\{G_n\}$ will be a fixed G. However, we formulate the results of this section in terms of sequences $\{G_n\}$ to allow extra generality. Note that if $G_n \equiv G$, $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$, then there always exists a $g \in L_r(G)$, $r = 1, 2$, such that $g > 0$, $0 < \int g^2 dG < \infty$.

Define, for $y \in \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^p$, $1 \leq j \leq p$,

$$S_j^o(y, \mathbf{u}) = V_{jd}(y, \mathbf{A}\mathbf{u}), \qquad Y_j^o(y, \mathbf{u}) := S_j^o(y, \mathbf{u}) - \mu_j^o(y, \mathbf{u}). \tag{13}$$

Note that for each j, $S_j^o$, $\mu_j^o$, $Y_j^o$ are the same as in (2.3.2) applied to $X_{ni} = Y_{ni}$, $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$ and $d_{ni} = d_{nij}$, $1 \leq i \leq n$, $1 \leq j \leq p$.

*Notation.* For any functions $g, h : \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$,

$$|g_{\mathbf{u}} - h_{\mathbf{v}}|_n^2 := \int \{g(y, \mathbf{u}) - h(y, \mathbf{v})\}^2 dG_n(y).$$

Occasionally we write $|g|_n^2$ for $|g_{\mathbf{0}}|_n^2$.

**Lemma 5.5.1.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective* d.f.'s $F_{n1}, \ldots, F_{nn}$. *Then* (5) *implies*

$$E\sum_{j=1}^p |Y_{j\mathbf{0}}^o|_n^2 = O(1). \tag{14}$$

**Proof.** By Fubini's Theorem,

$$E\sum_{j=1}^p |Y_{j\mathbf{0}}|_n^2 = \int \Sigma_i \, \|\mathbf{d}_i\|^2 \, F_i(1 - F_i) \, dG_n \tag{15}$$

and hence (5) implies the Lemma. □

**Lemma 5.5.2.** *Let* $\{Y_{ni}\}$ *be as in Lemma* 5.5.1. *Then assumptions* (1) – (4), (6) – (10) *imply that, for every* $0 < B < \infty$,

$$E \sup_{\|\mathbf{u}\| \leq B} \sum_{j=1}^p |Y_{j\mathbf{u}}^o - Y_{j\mathbf{0}}^o|_n^2 = o(1). \tag{16}$$

**Proof.** By Fubini's Theorem, $\forall\ \mathbf{u} \in \mathcal{N}(B)$,

$$\begin{aligned} E\sum_{j=1}^p |Y_{j\mathbf{u}}^o - Y_{j\mathbf{0}}^o|_n^2 &\leq \int \Sigma_i \, \|\mathbf{d}_i\|^2 |F_i(y + \mathbf{c}_i'\mathbf{u}) - F_i(y)| \, dG_n \\ &\leq \int_{-b_n}^{b_n} \Big( \int \gamma_n(y + x) \, dG_n(y) \Big) \, dx \end{aligned} \tag{17}$$

where $b_n = B \max_i \kappa_i$, $\gamma_n$ as in (6). Therefore, by assumption (6),

$$(18)\qquad E\sum_{j=1}^{p}|Y_{ju}^{o}-Y_{j0}^{o}|_n^2=o(1),\qquad \forall\ \mathbf{u}\in\mathbb{R}^p.$$

To complete the proof of (16), because of the compactness of $\mathcal{N}(B) := \{\mathbf{u}\in\mathbb{R}^p;\ \|\mathbf{u}\|\le B\}$, it suffices to show that $\forall\ \epsilon>0\ \exists$ a $\delta>0$ such that $\forall\ \mathbf{v}\in\mathcal{N}(B)$,

$$\limsup_n E\sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta}\sum_{j=1}^{p}|L_{j\mathbf{u}}-L_{j\mathbf{v}}|\le\epsilon,$$

where

$$L_{j\mathbf{u}} := |Y_{ju}^{o}-Y_{j0}^{o}|_n^2,\quad \mathbf{u}\in\mathbb{R}^p,\ 1\le j\le p.$$

Expand the quadratic, apply the C–S inequality to the cross product terms to obtain

$$(20)\quad |L_{j\mathbf{u}}-L_{j\mathbf{v}}|\le|Y_{ju}^{o}-Y_{jv}^{o}|_n^2+2|Y_{ju}^{o}-Y_{jv}^{o}|_n\,|Y_{jv}^{o}-Y_{j0}^{o}|_n,\ 1\le j\le p.$$

Moreover,

$$(21)\qquad |Y_{ju}^{o}-Y_{jv}^{o}|_n^2\le2\{|S_{ju}^{o}-S_{jv}^{o}|_n^2+|\mu_{ju}^{o}-\mu_{jv}^{o}|_n^2\},$$

$$|S_{ju}^{o}-S_{jv}^{o}|_n^2\le2\{|S_{ju}^{+}-S_{jv}^{+}|_n^2+|S_{ju}^{-}-S_{jv}^{-}|_n^2\},$$

$$|\mu_{ju}^{o}-\mu_{jv}^{o}|_n^2\le2\{|\mu_{ju}^{+}-\mu_{jv}^{+}|_n^2+|\mu_{ju}^{-}-\mu_{jv}^{-}|_n^2\},\qquad 1\le j\le p,$$

where $S_j^{\pm}$, $\mu_j^{\pm}$ are the $S_j^{o}$, $\mu_j^{o}$ with $d_{ij}$ replaced by $d_{ij}^{\pm}$, $d_{ij}^{+} := \max(0, d_{ij})$, $d_{ij}^{-} := d_{ij}^{+}-d_{ij}$, $1\le i\le n$, $1\le j\le p$.

Now, $\|\mathbf{u}-\mathbf{v}\|\le\delta$, nonnegativity of $\{d_{ij}^{\pm}\}$, and the monotonicity of $\{F_i\}$ yields (use (2.3.15) here), that for all $1\le j\le p$,

$$|\mu_{ju}^{\pm}-\mu_{jv}^{\pm}|_n^2\le\int[\Sigma_i\, d_{ij}^{\pm}\{F_i(y+\mathbf{c}_i'\mathbf{v}+\delta\kappa_i)-$$
$$-F_i(y+\mathbf{c}_i'\mathbf{v}-\delta\kappa_i)\}]^2\,dG_n(y).$$

Therefore, by assumption (7),

$$(22)\qquad \limsup_n\sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta}\sum_{j=1}^{p}|\mu_{ju}^{o}-\mu_{jv}^{o}|_n^2\le4k\delta^2.$$

By the monotonicity of $S_j^{\pm}$ and (2.3.15), $\|\mathbf{u}-\mathbf{v}\|\le\delta$ implies that for all $1\le j\le p$, $y\in\mathbb{R}$,

$$
\begin{aligned}
-\Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - \mathbf{v}'\mathbf{c}_i - y \le 0) \\
\le S_j^{\pm}(y, \mathbf{u}) - S_j^{\pm}(y, \mathbf{v}) \\
\le \Sigma_i d_{ij}^{\pm} I(0 < Y_i - \mathbf{v}'\mathbf{c}_i - y \le \delta\kappa_i).
\end{aligned}
$$

This in turn implies (using the fact that $a \le b \le c$ implies $b^2 \le a^2 + c^2$ for any reals $a, b, c$)

$$
\begin{aligned}
&\{S_j^{\pm}(y, \mathbf{u}) - S_j^{\pm}(y, \mathbf{v})\}^2 \\
&\quad \le \{\Sigma_i d_{ij}^{\pm} I(0 < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i)\}^2 + \\
&\qquad + \{\Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le 0)\}^2 \\
&\quad \le 2\,\{\Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i)\}^2
\end{aligned}
$$

for all $1 \le j \le p$ and all $y \in \mathscr{R}$. Now use the fact that for a, b real, $(a + b)^2 \le 2a^2 + 2b^2$ to conclude that, for all $1 \le j \le p$,

$$
\begin{aligned}
(23)\quad &|S_{j\mathbf{u}}^{\pm} - S_{j\mathbf{v}}^{\pm}|_n^2 \\
&\le 4\int \{\Sigma_i d_{ij}^{\pm}[I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i) - \\
&\qquad - p_i(y, \mathbf{v}, \delta)]\}^2\, dG_n(y) + \\
&\qquad + 4\int \{\Sigma_i d_{ij}^{\pm}\, p_i(y, \mathbf{v}, \delta)\}^2 dG_n(y) \\
&= 4\{I_j + II_j\}, \qquad\qquad \text{(say)},
\end{aligned}
$$

where $p_i(y, \mathbf{v}, \delta) \equiv F_i(y + \mathbf{v}'\mathbf{c}_i + \delta\kappa_i) - F_i(y + \mathbf{v}'\mathbf{c}_i - \delta\kappa_i)$.

But $(d_{ij}^{\pm})^2 \le d_{ij}^2$ for all i and j implies that

$$
\begin{aligned}
E\sum_{j=1}^{p} I_j &= \sum_{j=1}^{p} \int \Sigma_i (d_{ij}^{\pm})^2\, p_i(y, \mathbf{v}, \delta)\,(1 - p_i(y, \mathbf{v}, \delta))dG_n(y) \\
&\le \int \Sigma_i \|\mathbf{d}_i\|^2\, p_i(y, \mathbf{v}, \delta) dG_n(y) \le \int_{a_n}^{b_n} \left(\int \gamma_n(y + s) dG_n(y)\right) ds,
\end{aligned}
$$

by (3) and Fubini, where $a_n = (-B - \delta)\max_i \kappa_i$, $b_n = (B + \delta)\max_i \kappa_i$, and where $\gamma_n$ is defined in (6). Therefore, by the assumption (6),

$$
(24)\qquad E\sum_{j=1}^{p} I_j = o(1).
$$

From the definition of $II_j$ in (23) and the assumption (7),

$$\limsup_n \sum_{j=1}^{p} II_j \leq k\delta^2. \tag{25}$$

From (21) – (25), we obtain

$$\limsup_n E \sup_{\|\mathbf{u}-\mathbf{v}\|\leq\delta} \sum_{j=1}^{p} |Y^o_{j\mathbf{u}} - Y^o_{j\mathbf{v}}|^2_n \leq 40k\,\delta^2. \tag{26}$$

Thus if we choose $0 < \delta \leq (\epsilon/40k)^{1/2}$, then (19) will follow from (26), (20) and (18). This also completes the proof of (16). □

To state the next theorem we need

$$\hat{K}_{\mathbf{D}}(\mathbf{t}) := \sum_{j=1}^{p} \int \{Y^o_j(y, 0) + \mathbf{t}'\mathbf{R}_j(y) + m_j(y)\}^2\, dG_n(y). \tag{27}$$

In (28) below, the $G$ in $K_{\mathbf{D}}$ is assumed to have been replaced by the sequence $G_n$, just for extra generality.

**Theorem 5.5.1.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Suppose that* $\{X, F_{ni}, H_{ni}, \mathbf{D}, G_n\}$ *satisfy* (1) – (10). *Then, for every* $0 < B < \infty$,

$$E \sup_{\|\mathbf{u}\|\leq B} |K_{\mathbf{D}}(A\mathbf{u}) - \hat{K}_{\mathbf{D}}(A\mathbf{u})| = o(1). \tag{28}$$

**Proof.** Write $K, \hat{K}$ etc. for $K_{\mathbf{D}}, \hat{K}_{\mathbf{D}}$ etc. Note that

$$\begin{aligned} K(A\mathbf{u}) &= \sum_{j=1}^{p} \int [S^o_j(y, \mathbf{u}) - \mu^o_j(y) + m_j(y)]^2\, dG_n(y) \\ &= \sum_{j=1}^{p} \int [Y^o_j(y, \mathbf{u}) - Y^o_j(y) + Y^o_j(y) + \mathbf{u}'\boldsymbol{\nu}_j(y) + m_j(y) \\ &\qquad + \mu^o_j(y, \mathbf{u}) - \mu^o_j(y) - \mathbf{u}'\boldsymbol{\nu}_j(y)]^2\, dG_n(y) \end{aligned}$$

where $Y^o_j(y) \equiv Y^o_j(y, 0)$, $\mu^o_j(y) \equiv \mu^o_j(y, 0)$. Expand the quadratic and use the C–S inequality on the cross product terms to obtain

$$\begin{aligned} &|K(A\mathbf{u}) - \hat{K}(A\mathbf{u})| \\ &\leq \sum_{j=1}^{p} \Big\{ |Y^o_{j\mathbf{u}} - Y^o_{j\mathbf{v}}|^2_n + |\mu^o_{j\mathbf{u}} - \mu^o_j - \mathbf{u}'\boldsymbol{\nu}_j|^2_n \\ &\quad + 2|Y^o_{j\mathbf{u}} - Y^o_{j\mathbf{v}}|_n \big[|Y^o_j + \mathbf{u}'\boldsymbol{\nu}_j + m_j|_n + |\mu^o_{j\mathbf{u}} - \mu^o_j - \mathbf{u}'\boldsymbol{\nu}_j|_n\big] \\ &\quad + 2|Y^o_j + \mathbf{u}'\boldsymbol{\nu}_j + m_j|_n \cdot |\mu^o_{j\mathbf{u}} - \mu^o_j - \mathbf{u}'\boldsymbol{\nu}_j|_n \Big\}. \end{aligned} \tag{29}$$

In view of Lemmas 5.5.1, 5.5.2 and assumptons (8) and (10), (28) will follow from (29) if we prove

$$\sup_{\|\mathbf{u}\|\le B} \sum_{j=1}^{p} |\mu_{j\mathbf{u}}^{o} - \mu_j^{o} - \mathbf{u}'\nu_j|_n^2 = o(1). \tag{30}$$

Let $\xi_{j\mathbf{u}} := |\mu_{j\mathbf{u}}^{o} - \mu_j^{o} - \mathbf{u}'\nu_j|_n^2$, $1 \le j \le p$, $\mathbf{u}\in\mathbb{R}^p$. In view of the compactness of $\mathcal{N}(B)$ and the assumption (9), it suffices to prove that $\forall\ \epsilon > 0$, $\exists$ a $\delta > 0$ $\ni\ \forall\ \mathbf{v} \in \mathcal{N}(B)$,

$$\limsup_n \sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta} \sum_{j=1}^{p} |\xi_{j\mathbf{u}} - \xi_{j\mathbf{v}}| \le \epsilon. \tag{31}$$

But

$$\begin{aligned}|\xi_{j\mathbf{u}} - \xi_{j\mathbf{v}}| \le 2 \Big\{ & |\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n^2 + \|\mathbf{u}-\mathbf{v}\|^2\, \|\nu_j\|_n^2 \\ & + \xi_{j\mathbf{v}}^{1/2} \left[ |\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n + \|\mathbf{u}-\mathbf{v}\|\, \|\nu_j\|_n \right] \\ & + |\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n\, \|\mathbf{u}-\mathbf{v}\|\, \|\nu_j\|_n \Big\}.\end{aligned}$$

Hence, from (22) and the assumption (9),

$$\text{l.h.s. } (31) \le 2\,\{4k\delta^2 + \delta^2(a + 2k^{1/2}a^{1/2})\} = k_1\delta^2$$

where $a = \limsup_n \sum_{j=1}^{p} \|\nu_j\|_n^2$. Therefore, choose $\delta^2 \le \epsilon/k_1$ to obtain (31), hence (30) and therefore the Theorem. □

Our next goal is to obtain an analogue of (28) for $K_{\mathbf{D}}^{+}$. Before stating it rigorously, it helps to rewrite $K_{\mathbf{D}}^{+}$ in terms of standardized processes $\{Y_j^{o}\}$ and $\{\mu_j^{o}\}$ defined at (9). In fact, we have

$$\begin{aligned}K_{\mathbf{D}}^{+}(A\mathbf{u}) &= \sum_{j=1}^{p} \int [S_j^{o}(y,\mathbf{u}) - \Sigma_i d_{ij} + S_j^{o}(-y,\mathbf{u})]^2\, dG_n(y) \\ &= \sum_{j=1}^{p} \int [Y_j^{o}(y,\mathbf{u}) - Y_j^{o}(y) + Y_j^{o}(-y,\mathbf{u}) - Y_j^{o}(-y) \\ &\qquad + \mu_j^{o}(y,\mathbf{u}) - \mu_j^{o}(y) - \mathbf{u}'\nu_j(y) \\ &\qquad + \mu_j^{o}(-y,\mathbf{u}) - \mu_j^{o}(-y) - \mathbf{u}'\nu_j(y) \\ &\qquad + \mathbf{u}'\nu_j^{+}(y) + W_j^{+}(y) + m_j^{+}(y)]^2\, dG_n(y)\end{aligned}$$

where

$$W_j^+(y) := Y_j^o(y) + Y_j^o(-y), \qquad \nu_j^+(y) = \nu_j(y) + \nu_j(-y),$$

$$m_j^+(y) := \Sigma_i\, d_{ij}\, \{F_i(y) - 1 + F_i(-y)\}$$

$$= \mu_j^o(y) + \mu_j^o(-y) - \Sigma_i\, d_{ij}, \qquad y \in \mathbb{R},\ 1 \le j \le p.$$

Let

$$\text{(32)} \qquad \hat{K}_D^+(A\mathbf{u}) = \sum_{j=1}^{p} \int [W_j^+ + m_j^+ + \mathbf{u}'\nu_j^+]^2\, dG_n, \qquad \mathbf{u} \in \mathbb{R}^p.$$

Now proceeding as in (29), one obtains a similar upper bound for $|K_D^+(A\mathbf{u}) - \hat{K}_D^+(A\mathbf{u})|$ innvolving terms like those in r.h.s. of (29) and the terms like $|Y_{j\mathbf{u}}^o - Y_j^o|_{-n}$, $|\mu_{j\mathbf{u}}^o - \mu_j^o - \mathbf{u}'\nu_j|_{-n}$ $\|\nu_j\|_{-n}$, $|Y_j|_{-n}$, where for any function $h\colon \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$, $|h_{\mathbf{u}}|_{-n}^2 := \int h^2(-y, \mathbf{u})\, dG_n(y)$. It thus becomes apparent that one needs an analogue of Lemmas 5.5.1 and 5.5.2 with $G_n(\cdot)$ replaced by $G_n(-\cdot)$. That is, if the conditions (5) – (10) are also assumed to hold for measures $\{G_n(-\cdot)\}$ then obviously analogues of these lemmas will hold. Alternatively, the statement of the following theorem and the details of its proof are considerably simplified if one assumes $G_n$ to be symmetric around zero, as we shall do for convenience. Before stating the theorem, we state

**Lemma 5.5.3.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Assume* (1) – (4), (6), (7) *hold,* $\{G_n\}$ *satisfies* (5.3.8) *and that* (33) *hold, where*

$$\text{(33)} \qquad \int \Sigma_i \|d_{ni}\|^2 \{F_{ni}(-y) + 1 - F_{ni}(y)\}\, dG_n(y) = O(1).$$

*Then,*

$$\text{(34a)} \qquad E \sum_{j=1}^{p} |Y_{j0}^o|_{-n}^2 = O(1),$$

*and*

$$\text{(34b)} \qquad E \sup_{\|\mathbf{u}\| \le B} \sum_{j=1}^{p} |Y_{j\mathbf{u}}^o - Y_{j0}^o|_{-n}^2 = o(1), \qquad \forall\ 0 < B < \infty. \ \square$$

This lemma follows from Lemmas 5.5.1 and 5.5.2 because under (5.3.8), l.h.s.'s of (34a) and (34b) are equal to those of (14) and (16), respectively. The proof of the following theorem is similar to that of Theorem 5.5.1.

**Theorem 5.5.2.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Suppose that* $\{X, F_{ni}, D, G_n\}$ *satisfy* (1) –

(4), (6) – (9), (5.3.8) *for all* $n \geq 1$, (33) *and that*

$$\sum_{j=1}^{p} \int \{m_j^+(y)\}^2 \, dG_n(y) = O(1), \tag{35}$$

*Then,* $\forall \ 0 < B < \infty$,

$$E \sup_{\|u\| \leq B} |K_D^+(Au) - \hat{K}_D^+(Au)| = o(1). \tag{36}$$ □

**Remark 5.5.1.** Recall that we are interested in the asymptotic distribution of $A^{-1}(\hat{\beta}_D - \beta)$ which is a minimizer of $\hat{K}_D(\beta + Au)$ w.r.t. $u$. Since $\hat{\beta}_D$ satisfies (5.3.22), there is no loss of generality in taking the true $\beta$ equal to 0. Then (28) asserts that $(1/2)\hat{K}_D$ satisfies (5.4.A1) with

$$\theta_0 = 0, \ \delta_n = A^{-1}, \ S_n = A^{-1}\mathcal{S}_n, \ W_n = A\,\mathcal{B}_n\,A, \tag{37}$$

$$\mathcal{S}_n := -\int \Gamma_n(y)\{Y_D^o(y) + m_D(y)\}\, dG_n(y),$$

$$\mathcal{B}_n := \int \Gamma_n(y)\Gamma_n'(y)\, dG_n(y),$$

where $\Gamma_n(y) = AX'\Lambda^*(y)D$, $\Lambda^*$ as in (4.2.1), $Y_D^{o\prime} := (Y_1^o, ...., Y_p^o)$ and $m_D' := (m_1, ...., m_p)$.

In view of Lemma 5.5.1, the assumptions (5) and (10) imply that $EK_D(0) = O(1)$, thereby ensuring the validity of (5.4.A4).

Similarly, (36) asserts that $(1/2)K_D^+$ satisfies (5.4.A1) with

$$\theta_0 = 0, \ \delta_n = A^{-1}, \ S_n = A^{-1}\mathcal{S}_n^+, \ W_n = A\mathcal{B}_n^+ A, \tag{38}$$

$$\mathcal{S}_n^+ := -\int \Gamma_n^+(y)\{W_D^+(y) + m_D^+(y)\}\, dG_n(y),$$

$$\mathcal{B}_n^+ := \int \Gamma_n^+(y)\Gamma_n^{+\prime}(y)\, dG_n(y),$$

where $\Gamma_n^+(y) := AX'\Lambda^+(y)D$, $\Lambda^+(y) := \Lambda^*(y) + \Lambda^*(-y)$, $y \in \mathbb{R}^p$, $W_D^{+\prime} := (W_1^+, ..., W_p^+)$ and $m_D^{+\prime} := (m_1^+, ..., m_p^+)$.

In view of (12), (31), (33) and (5.3.8) it follows that (5.4.A4) is satisfied by $K_D^+(0)$.

Theorem 5.4.1 enables one to study the asymptotic distribution of $\hat{\beta}_D$ when in (1.1.1) the actual error d.f. $F_{ni}$ is not necessarily equal to the modeled d.f. $H_{ni}$, $1 \le i \le n$. Theorem 5.4.2 enables one to study the asymptotic distribution of $\beta_D^+$ when in (1.1.1) the error d.f. $F_{ni}$ is not necessarily symmetric around $\mathbf{0}$, but we model it to be so, $1 \le i \le n$. □

So far we have not used the assumptions (11) and (12). They will be now used to obtain (5.4.A5) for $K_D$ and $K_D^+$.

**Lemma 5.5.4.** *In addition to the assumptions of Theorem* 5.5.1 *assume that* (11) *and* (12) *hold. Then,* $\forall\ \epsilon > 0,\ 0 < z < \infty,\ \exists$ N (*depending only on* $\epsilon$) *and a* B (*depending on* $\epsilon$, z) $\ni\ 0 < B < \infty$,

$$P\Big(\inf_{\|u\|>B} K_D(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N, \tag{39}$$

$$P\Big(\inf_{\|u\|>B} \hat{K}_D(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N. \tag{40}$$

**Proof.** As usual write K, $\hat{K}$ etc. for $K_D$, $\hat{K}_D$ etc. Recall the definition of $\bar{\Gamma}_n$ from (11). Let $k_n(\theta) := \theta'\bar{\Gamma}_n\theta$, $\theta \in \mathscr{R}^p$. By the C–S inequality and (11),

$$\sup_{\|\theta\|=1} |k_n(\theta)|^2 \le \|\bar{\Gamma}_n\|^2 \le \sum_{j=1}^{p} \|\nu_j\|_n^2 \cdot |g_n|_n^2 = O(1). \tag{41}$$

Fix an $\epsilon > 0$ and a $z \in (0, \infty)$. Define, for $\mathbf{t} \in \mathbb{R}^p$, $1 \le j \le p$,

$$\hat{V}_j(\mathbf{t}) := \int \{Y_j^o + \mathbf{t}'\mathbf{R}_j + m_j\}\, g_n\, dG_n,$$

$$V_j(\mathbf{t}) := \int [V_{jd}(y, \mathbf{t}) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]\, g_n(y)\, dG_n(y).$$

Also, let $\hat{V}' := (\hat{V}_1, \ldots, \hat{V}_p)$, $V' := (V_1, \ldots, V_p)$, $\gamma_n := |g_n|_n^2$, $\gamma := \limsup_n \gamma_n$.

Write a $\mathbf{u} \in \mathbb{R}^p$ with $\|\mathbf{u}\| > B$ as $\mathbf{u} = r\theta$, $|r| > B$, $\|\theta\| = 1$. Then, by the C–S inequality,

$$\inf_{\|u\|>B} K(Au) \ge \inf_{|r|>B,\,\|\theta\|=1} (\theta' V(rA\theta))^2/\gamma_n,$$

$$\inf_{\|u\|>B} \hat{K}(Au) \ge \inf_{|r|>B,\,\|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n.$$

It thus suffices to show that $\exists$ a $B \in (0, \infty)$ and N $\ni$

$$(\tilde{39})\qquad P\Big(\inf_{|r|>B,\,\|\theta\|=1} (\theta' V(rA\theta))^2/\gamma_n \geq z\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N,$$

$$(\tilde{40})\qquad P\Big(\inf_{|r|>B,\,\|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n \geq z\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N.$$

But, $\forall\ u \in \mathbb{R}^p$,

$$\|V(Au) - \hat{V}(Au)\| \leq 2\gamma_n \sum_{j=1}^{p} \Big\{ |Y_{ju}^o - Y_{jv}^o|_n^2 + |\mu_{ju}^o - \mu_j^o - u'\nu_j|_n^2 \Big\}.$$

Thus, from (11), (16) and (30), it follows that $\forall\ B \in (0, \infty)$,

$$(42)\qquad \sup_{\|u\| \leq B} \|V(Au) - \hat{V}(Au)\| = o_p(1).$$

Now rewrite

$$\theta' \hat{V}(rA\theta) = \theta' T + r\, k_n(\theta), \qquad T' := (T_1, \ldots, T_p) \quad \text{with}$$

$$T_j := \int \{Y_j^o + m_j\}\, g_n\, dG_n, \qquad 1 \leq j \leq p.$$

Again, by the C–S inequality, Fubini, (16) and the assumptions (10) and (11) it follows that $\exists\ N_1$ and $b$, possibly both depending on $\epsilon$, such that

$$(43)\qquad P(\|T\| \leq b) \geq 1 - (\epsilon/2), \qquad n \geq N_1.$$

Now choose $B$ such that

$$(44)\qquad B \geq (b + (za)^{1/2})\, \alpha^{-1},$$

where $\alpha$ is as in (11). Then, with $\alpha_n := \inf\{|k_n(\theta)|;\ \|\theta\| = 1\}$,

$$(45)\qquad P\Big(\inf_{|r|=B,\,\|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n \geq z\Big)$$

$$= P(|\theta' \hat{V}(rA\theta)| \geq (z\gamma_n)^{1/2},\ \forall\ \|\theta\| = 1,\ |r| = B)$$

$$\geq P(|\,|\theta' T| - |r|\ |k_n(\theta)|\,| \geq (z\gamma_n)^{1/2},\ \forall\ \|\theta\| = 1,\ |r| = B)$$

$$\geq P(\|T\| \leq -(z\gamma_n)^{1/2} + B\,\alpha_n) \geq P(\|T\| \leq -(z\gamma)^{1/2} + B\,\alpha)$$

$$\geq P(\|T\| \leq b) \geq 1 - (\epsilon/2), \qquad \forall\ n \geq N_1.$$

In the above, the first inequality follows from the fact that $||d| - |c|| \le |d + c|$, d, c reals; the second uses the fact that $|\boldsymbol{\theta}' \boldsymbol{T}| \le \|\boldsymbol{T}\|$ for all $\|\boldsymbol{\theta}\| = 1$; the third uses the relation $(-\infty, -(z\gamma)^{1/2} + B\,\alpha) \subset (-\infty, -(z\gamma_n)^{1/2} + B\,\alpha_n)$; while the last inequality follows from (43) and (44).

Observe that $\boldsymbol{\theta}' \hat{\boldsymbol{V}}(r\mathbf{A}\boldsymbol{\theta})$ is monotonic in r for every $\|\boldsymbol{\theta}\| = 1$. Therefore, (45) implies ($\tilde{40}$) and hence (40) in a straight forward fashion.

Next, consider $\boldsymbol{\theta}' \boldsymbol{V}(r\mathbf{A}\boldsymbol{\theta})$. Rewrite

$$\boldsymbol{\theta}' \boldsymbol{V}(r\mathbf{A}\boldsymbol{\theta}) = \int \sum_{i=1}^{n} (\boldsymbol{\theta}' \mathbf{d}_{ij})[I(Y_{ni} \le y + r\mathbf{x}'_{ni}\mathbf{A}\boldsymbol{\theta})) - H_{ni}(y)]\, g_n(y)\, dG_n(y)$$

which, in view of the assumption (12), shows that $\boldsymbol{\theta}' \boldsymbol{V}(r\mathbf{A}\boldsymbol{\theta})$ is monotonic in r for every $\|\boldsymbol{\theta}\| = 1$. Therefore, by (42) $\exists$ $N_2$, depending on $\epsilon$, $\ni$

$$\begin{aligned} P\Big(\inf_{|r|>B,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \boldsymbol{V}(r\mathbf{A}\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) & \\ \ge P\Big(\inf_{|r|=B,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \boldsymbol{V}(r\mathbf{A}\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) & \\ \ge P\Big(\inf_{|r|=B,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \hat{\boldsymbol{V}}(r\mathbf{A}\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) - (\epsilon/2), & \qquad \forall\, n \ge N_2, \\ \ge 1 - \epsilon, & \qquad \forall\, n \ge N_2 \vee N_1, \end{aligned}$$

by (45). This proves ($\tilde{39}$) and hence (39). □

The next lemma gives an analogue of the previous lemma for $K^+_{\mathbf{D}}$. Since the proof is quite similar no details will be given.

**Lemma 5.5.5.** *In addition to the assumptions of Theorem* 5.5.2 *assume that* ($11^+$) *and* (12) *hold, where* ($11^+$) *is the condition* (11) *with* $\Gamma_n$ *replaced by* $\Gamma_n^+ := (\nu_1^+, \ldots, \nu_p^+)$ *and where* $\{\nu_j^+\}$ *are defined just above* (32).

*Then,* $\forall\ \epsilon > 0$, $0 < z < \infty$, $\exists$ N *(depending only on* $\epsilon$*) and a* B *(depending on* $\epsilon$, z*)* $\ni$

$$\text{(46)} \qquad P\Big(\inf_{\|\mathbf{u}\|>B} K^+_{\mathbf{D}}(\mathbf{A}\mathbf{u}) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N,$$

$$\text{(47)} \qquad P\Big(\inf_{\|\mathbf{u}\|>B} \hat{K}^+_{\mathbf{D}}(\mathbf{A}\mathbf{u}) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N. \ □$$

The above two lemmas verify (5.4.A5) for the two dispersions K and $K^+$. Also note that (40) together with (5) and (10) imply that $\|A^{-1}(\hat{\Delta} - \beta)\| = O_p(1)$, where $\hat{\Delta}$ is defined at (49) below. Similarly, (47), (5), (35) and the symmetry assumption (5.3.8) about $\{G_n\}$ imply that $\|A^{-1}(\Delta^+ - \beta)\| = O_p(1)$, where $\Delta^+$ is defined at (53) below. The proofs of these facts are exactly similar to that of (5.4.2) given in the proof of Theorem 5.4.1.

In view of Remark 5.5.1 and Theorem 5.4.1, we have now proved the following theorems.

**Theorem 5.5.3.** *Assume that* (1.1.1) *holds with the modeled and actual d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{H_{ni}, 1 \le i \le n\}$ *and* $\{F_{ni}, 1 \le i \le n\}$, *respectively. In addition, suppose that* (1) – (12) *hold. Then*

(48) $$(\hat{\beta}_D - \hat{\Delta})' A^{-1} \mathscr{B}_n A^{-1}(\hat{\beta}_D - \hat{\Delta}) = o_p(1),$$

*where* $\hat{\Delta}$ *satisfies the equation*

(49) $$\mathscr{B}_n A^{-1}(\hat{\Delta} - \beta) = \mathscr{T}_n.$$

*If, in addition,*

(50) $$\mathscr{B}_n^{-1} \text{ exists for } n \ge p,$$

*then,*

(51) $$A^{-1}(\hat{\beta}_D - \beta) = \mathscr{B}_n^{-1} \mathscr{T}_n + o_p(1),$$

*where* $\mathscr{T}_n$ *and* $\mathscr{B}_n$ *are defined at* (37). □

**Theorem 5.5.4.** *Assume that* (1.1.1) *holds with the actual d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$. *In addition, suppose that* $\{X, F_{ni}, D, G_n\}$ *satisfy* (1)–(4), (6) – (9), (5.3.8) *for all* $n \ge 1$, (11), (12) *and* (33). *Then,*

(52) $$(\beta_D^+ - \Delta^+)' A^{-1} \mathscr{B}_n^+ A^{-1}(\beta_D^+ - \Delta^+) = o_p(1),$$

*where* $\Delta^+$ *satisfies the equation*

(53) $$\mathscr{B}_n^+ A^{-1}(\Delta^+ - \beta) = \mathscr{T}_n^+.$$

*If, in addition,*

(54) $$(\mathscr{B}_n^+)^{-1} \text{ exists for } n \ge p,$$

*then,*

(55) $$A^{-1}(\beta_D^+ - \beta) = (\mathscr{B}_n^+)^{-1}\mathscr{T}_n^+ + o_p(1),$$

*where* $\mathscr{T}_n^+$ *and* $\mathscr{B}_n^+$ *are defined at* (38). □

**Remark 5.5.2.** If $\{F_i\}$ are symmetric about zero then $m_D^+ \equiv 0$ and $\beta_D^+$ is consistent for $\beta$ even if the errors are not identically distributed. On the other hand, if the errors are identically distributed, but not symmetrically, then $\beta_D^+$ will be asymptotically biased. This is not surprising because here the symmetry, rather than the identically distributed nature of the errors is relevant.

If $\{F_i\}$ are symmetric about an unknown common point then that point can be also estimated by the above m.d. method by simply augmenting the design matrix to include the column 1, if not present already. □

Next we turn to the $K_D^*$ and $\beta_D^*$ (5.2.22) and (5.2.23). First we state a theorem giving an analogue of (28) for $K_D^*$. Let $Y_j$, $\mu_j$ be $Y_d$, $\mu_d$ of (2.3.1) with $\{d_{ni}\}$ replaced by $\{d_{nij}\}$, $j = 1, \ldots, p$, $X_{ni}$ replaced by $Y_{ni}$ and $c_{ni} = A_1(x_{ni} - \bar{x}_n)$, $1 \le i \le n$, where $A_1$ and $\bar{x}_n$ are defined at (4.3.11). Set

(56) $$R_j^*(s) := \Sigma_i\, (d_{nij} - \tilde{d}_{nj}(s))(x_{ni} - \bar{x}_n)\, q_{ni}(s),$$

where, for $1 \le j \le p$, $\tilde{d}_{nj}(s) := n^{-1}\,\Sigma_i\, d_{nij}\,\ell_{ni}(s)$, $0 \le s \le 1$, with $\{\ell_{ni}\}$ as in (3.2.35) and $q_{ni} \equiv f_{ni}(H^{-1})$, $1 \le i \le n$. Let

(57) $$\hat{K}_D^*(t) := \sum_{j=1}^{p} \int_0^1 \{Y_j(s,0) - t' R_j^*(s) + \mu_j(s,0)\}^2\, dL_n(s).$$

In (59) below, L in $K_D^*$ is supposed to have been replaced by $L_n$.

**Theorem 5.5.5.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Assume* $\{D, X, F_{ni}\}$ *satisfy* (1), (2), (3), (2.3.3b), (3.2.12), (3.2.35) *and* (3.2.36) *with* $w_i = d_{ij}$, $1 \le j \le p$, $1 \le i \le n$. *Let* $\{L_n\}$ *be a sequence of d.f.'s on* [0, 1] *and assume that*

(58) $$\sum_{j=1}^{p} \int_0^1 \mu_j^2(s,0)\, dL_n(s) = O(1).$$

*Then, for every* $0 < B < \infty$,

(60) $$\sup_{\|u\| \le B} |K_D^*(Au) - \hat{K}_D^*(Au)| = o_p(1).$$

**Proof.** The proof of (60) uses the a.u.l. result of Theorems 3.2.1 and 3.2.4. Details are left out as an exercise. □

The result (60) shows that the dispersion $K_D^*$ satisfies (5.4.A1) with

$$\theta_0 = 0,\ \delta_n = A_1^{-1},\ S_n = A_1^{-1}\mathcal{S}_n^*,\ W_n = A_1\mathcal{B}_n^* A_1, \tag{61}$$

$$\mathcal{S}_n^* := -\int_0^1 \Gamma_n^*(s)\{Y_D(s) + \mu_D(s)\}\, dL_n(s),$$

$$\mathcal{B}_n^* := \int_0^1 \Gamma_n^*(s)\Gamma_n^{*\prime}(s)\, dL_n(s),$$

where $\Gamma_n^*(s) = A_1' X_c \Lambda(s) D(s)$, $D(s) := ((d_{nij} - \tilde{d}_{nj}(s)),\ 1 \le i \le n,\ 1 \le j \le p;\ \Lambda(s)$ as in (2.3.32), $0 \le s \le 1$; $X_c$ as in (4.2.11); $Y_D := (Y_1, \ldots, Y_p)'$, $\mu_D' := (\mu_1, \ldots, \mu_p)$ *with* $Y_j(s) \equiv Y_j(s, 0)$, $\mu_j(s) \equiv \mu_j(s, 0)$.

Call the condition (11) by the name of (11*) if it holds when $(\Gamma_n, G_n)$ is replaced by $(\Gamma_n^*, L_n)$. Analogous to Theorem 5.5.4 we have

**Theorem 5.5.6.** *Assume that* (1.1.1) *holds with the actual d.f.'s of the errors* $\{e_{ni},\ 1 \le i \le n\}$ *equal to* $\{F_{ni},\ 1 \le i \le n\}$. *In addition, assume that* $\{D, X, F_{ni}\}$ *satisfy* $(NX^*)$, (2), (3), (2.3.3b), (3.2.12), (3.2.35), (3.2.36) *with* $w_i = d_{ij}$, $1 \le j \le p$, $1 \le i \le n$, (11*) *and* (12). *Let* $\{L_n\}$ *be a sequence of d.f.'s on* $[0, 1]$ *satisfying* (58). *Then*

$$(\beta_D^* - \Delta^*)' A^{-1} \mathcal{B}_n^* A^{-1} (\beta_D^* - \Delta^*) = o_p(1), \tag{62}$$

*where* $\Delta^*$ *satisfies the equation*

$$\mathcal{B}_n^* A^{-1}(\Delta^* - \beta) = \mathcal{S}_n^*. \tag{63}$$

*If, in addition,*

$$(\mathcal{B}_n^*)^{-1} \text{ exists for } n \ge p, \tag{64}$$

*then,*

$$A^{-1}(\beta_D^* - \beta) = (\mathcal{B}_n^*)^{-1}\mathcal{S}_n^* + o_p(1), \tag{65}$$

*where* $\mathcal{S}_n^*$ *and* $\mathcal{B}_n^*$ are *defined at* (61).

The proof of this theorem is similar to that of Theorem 5.5.3. The details are left out for interested readers. See also Section 4.3. □

**Remark 5.5.3.** ***Discussion of the assumptions*** (1) – (10). Among the assumptions (1) – (10), the assumptions (7) and (9) are relatively harder to verify. First, we shall give some sufficient conditions that will imply (7), (9) and the other assumptions. Then, we shall discuss these assumptions in detail for three cases, v.i.z., the case when the errors are correctly modeled to be i.i.d. F, F a known d.f., the case when we model the errors to be i.i.d. F but they actually have heteroscedastic gross errors distributions, and finally, the case when the errors are modeled to be i.i.d. F but they actually are heteroscedastic due to difference in scales.

To begin with consider the following assumptions.

(66) *For any sequences of numbers* $\{a_{ni}, b_{ni}\}$, $a_{ni} < b_{ni}$, $\max_i (b_{ni} - a_{ni}) \longrightarrow 0$,

$$\limsup_n \max_i (b_{ni}-a_{ni})^{-1} \int_{a_{ni}}^{b_{ni}} \int \{f_{ni}(y+z)-f_{ni}(y)\}^2 \, dG_n(y)\, dz = 0.$$

(67)
$$\max_i \int f_{ni}^2 \, dG_n = O(1).$$

*Claim* 5.5.1. *Assumptions* (1) – (4), (66), (67) *imply* (7) *and* (9).

Proof. Use the C–S inequality twice, the fact that $(d_{ij}^{\pm})^2 \le d_{ij}^2$ for all i, j, and (2) to obtain

$$\sum_{j=1}^{p} \int [\sum_i d_{ij}^{\pm} \{F_i(y + c_i'\mathbf{v} + \delta\kappa_i) - F_i(y + c_i'\mathbf{v} - \delta\kappa_i)\}]^2 \, dG_n(y)$$

$$\le 2 \,\Sigma_i \|d_i\|^2 \int \Sigma_i \,\delta\kappa_i \int_{a_i}^{b_i} f_i^2 (y + z)\, dz \, dG_n(y)$$

$$\le 4p^2\delta^2 \max_i (2\delta\kappa_i)^{-1} \int_{a_i}^{b_i} \int f_i^2 (y+z)\, dG_n(y)dz, \quad \text{(by Fubini)},$$

where $a_i = -\kappa_i\delta + c_i'\mathbf{v}$, $b_i = \kappa_i\delta + c_i'\mathbf{v}$, $1 \le i \le n$. Therefore, by (66), (67) and (1),

$$\text{l.h.s. } (7) \le 4p^2\delta^2 k, \qquad (k = \limsup_n \max_i |f_i|_n^2),$$

which shows that (7) holds.

Next, by (2) and two applications of the C–S inequality

$$\text{l.h.s. } (9) = \sum_{j=1}^{p} \int [\Sigma_i \, d_{ij} \{F_i(y + c_i'\mathbf{u}) - F_i(y) - c_i'\mathbf{u} f_i(y)\}]^2 \, dG_n(y)$$

$$\le p \int \Sigma_i \{F_i(y + c_i'\mathbf{u}) - F_i(y) - c_i'\mathbf{u} f_i(y)\}^2 \, dG_n(y)$$

$$= p\Big\{\int \Sigma_i^+ \Big[\int_0^{c_i'u} (f_i(y+z) - f_i(y))dz\Big]^2 dG_n(y)$$

$$+ \int \Sigma_i^- \Big[-\int_{c_i'u}^{0} (f_i(y+z) - f_i(y))dz\Big]^2 dG_n(y)\Big\}$$

$$\leq 2p\,\Big\{\int [\Sigma_i^+ c_i'u \int_0^{c_i'u} \{f_i(y+z) - f_i(y)\}^2 dz$$

$$+ \Sigma_i^- (-c_i'u) \int_{c_i'u}^{0} \int \{f_i(y+z) - f_i(y)\}^2 dz]\, dG_n(y)$$

$$\leq [\max_i (2|c_i'u|)^{-1} \int_{-|c_i'u|}^{|c_i'u|} \int \{f_i(y+z) - f_i(y)\}^2 dG_n(y)dz] \cdot$$

$$\cdot 4p\, \Sigma_i (c_i'u)^2,$$

where $\Sigma_i^+$ $(\Sigma_i^-)$ is the sum over those $i$ for which $c_i'u \geq 0$ $(c_i'u < 0)$. Since $\Sigma_i (c_i'u)^2 \leq pB$ for all $u \in \mathcal{N}(B)$, (9) now follows from (66) and (1). □

Now we consider the three special cases mentioned above.

***Case* 5.5.1.** *Correctly modeled i.i.d. errors*: $F_{ni} \equiv F \equiv H_{ni}$, $G_n \equiv G$. Suppose that $F$ has a density $f$ w.r.t. $\lambda$. Assume that

(68) (a) $0 < \int f\, dG < \infty$, (b) $0 < \int f^2\, dG < \infty$.

(69) $\int F(1-F)\, dG < \infty$.

(70) (a) $\lim_{z\to 0} \int f(y+z)\, dG(y) = \int f\, dG$

(b) $\lim_{z\to 0} \int f^2(y+z)\, dG(y) = \int f^2\, dG$.

***Claim* 5.5.2:** *Assumptions* (1), (2), (4) *with* $G_n \equiv G$, (68) – (70) *imply* (1) – (10) *with* $G_n \equiv G$.

This is easy to see. In fact here (5) and (6) are equivalent to (68a), (69) and (70a); (66) and (67) are equivalent to (68b) and (70b). The LHS (10) = 0.

Note that if $G$ is absolutely continuous then (68) implies (70). If $G$ is purely discrete and $f$ continuous at the points of jumps of $G$ then (70) holds. In particular if $G = \delta_0$, i.e., if $G$ is degenerate at 0, $\infty > f(0) > 0$ and $f$ is continuous at 0 then (68), (70) are trivially satisfied. If $G(y) \equiv y$, (68a) and (70a) are *a priori* satisfied while (69) is equivalent to assuming that $E|e_1 - e_2| < \infty$, $e_1, e_2$ i.i.d. $F$.

If $dG(y) = \{F(y)(1 - F(y))\}^{-1} dF(y)$, the so called Darling–Anderson measure, then (68) – (70) are satisfied by a class of d.f.'s that includes normal, logistic and double exponential distributions.

*Case* 5.5.2. *Heteroscedastic gross errors*: $H_{ni} \equiv F$, $F_{ni} \equiv (1-\delta_{ni})F+\delta_{ni} F_0$. We shall also assume that $G_n \equiv G$. Let $f$ and $f_0$ be continuous densities of $F$ and $F_0$. Then $\{F_{ni}\}$ have densities $f_{ni} = f + \delta_{ni}(f_0 - f)$, $1 \le i \le n$. Hence (3) is satisfied. Consider the assumption

$$0 \le \delta_{ni} \le 1, \qquad \max_i \delta_{ni} \longrightarrow 0, \tag{71}$$

$$\int |F_0 - F|\, dG < \infty. \tag{72}$$

*Claim* 5.5.3. *Suppose that* $f_0$ *and* $f$ *satisfy* (68) *and* (70), $F$ *satisfies* (69), *and suppose that* (1), (2) *and* (4) *hold. Then* (71) *and* (72) *imply* (5) – (9).

**Proof.** The relation $f_i \equiv f + \delta_i(f_0 - f)$ implies that

$$\nu_j - \Sigma_i\, d_{ij}\, c_i\, f = \Sigma_i\, d_{ij}\, c_i\, \delta_i\, (f_0 - f), \qquad 1 \le j \le p,$$

and

$$\gamma_n - \Sigma_i\, \|d_i\|^2 f = \Sigma_i\, \|d_i\|^2\, \delta_i\, (f_0 - f).$$

Because $\Sigma_i \|d_i\|^2 \le p$, $\Sigma_i \|c_i\|^2 = p$, we obtain

$$\Big|\int [\gamma_n(y+x) - \Sigma_i \|d_i\|^2 f(y + x)]dG(y)\Big|$$
$$\le p \max_i \delta_i \Big|\int [f_0(y + x) - f(y + x)]dG(y)\Big|, \qquad \forall\ x \in \mathbb{R}.$$

Therefore, by (71), (68a) and (70a), it follows that (6) is satisfied. Similarly, the inequality

$$\sum_{j=1}^{p} \int \|\nu_j - \Sigma_i\, d_{ij}\, c_i\, f\|^2\, dG \le 2p^2 \max_i \delta_i^2 \{\int f_0^2\, dG + \int f^2\, dG\}$$

ensures the satisfaction of (8). The inequality

$$\Big|\int \Sigma_i \|d_i\|^2\{F_i(1 - F_i) - F(1 - F)\}\, dG\Big| \le 2p \max_i \delta_i \int |F_0 - F|\, dG,$$

(69), (71) and (72) imply (5). Next,

$$\int \{f_i(y+x) - f_i(y)\}^2 \, dG(y)$$
$$\leq 2(1+2\delta_i^2) \int \{f(y+x) - f(y)\}^2 \, dG(y) + 4\delta_i^2 \int \{f_0(y+x) - f_0(y)\}^2 \, dG(y).$$

Note that (68b), (70b) and the continuity of f imply that

$$\lim_{x \to 0} \int \{f(y+x) - f(y)\}^2 \, dG(y) = 0$$

and a similar result for $f_0$. Therefore from the above inequality, (70) and (71) we see that (66) and (67) are satisfied. By Claim 5.5.1, it follows that (7) and (9) are satisfied. □

Suppose that G is a *finite measure.* Then (F1) implies (68) – (70) and (72). In particular these assumptions are satisfied by all those f's that have finite Fisher information.

The assumption (10), in view of (72), amounts to requiring that

(73) $$\sum_{j=1}^{p} (\sum_i d_{ij} \, \delta_i)^2 = O(1).$$

But

(74) $$\sum_{j=1}^{p} (\sum_i d_{ij} \delta_i)^2 = \sum_{i=1}^{n} \sum_{k=1}^{n} d_i' \, \delta_i \, d_k \, \delta_k \leq (\sum_i \|d_i\| \, \delta_i)^2.$$

This and (2) suggest a choice of $\delta_i \equiv p^{-1/2} \|d_i\|$ will satisfy (73). Note that if $D = XA$ then $\|d_i\|^2 \equiv \mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i$.

When studying the robustness of $\hat{\beta}_X$ in the following section, $\delta_i^2 \equiv p^{-1}\mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i$ is a natural choice to use. It is an analogue of $n^{-1/2}$–contamination in the i.i.d. setup. □

***Case*** **5.5.3.** *Heteroscedastic scale errors:* $H_{ni} \equiv F$, $F_{ni}(y) \equiv F(\tau_{ni} y)$, $G_n \equiv G$. Let F have continuous density f. Consider the conditions

(75) $$\tau_{ni} \equiv \sigma_{ni} + 1; \quad \sigma_{ni} > 0, \ 1 \leq i \leq n; \quad \max_i \sigma_{ni} \longrightarrow 0.$$

(76) $$\lim_{s \to 1} \int |y|^j f^k(sy) \, dG(y) = \int |y|^j f^k(y) \, dG(y), \ j = 1, k = 1;$$
$$j = 0, k = 1, 2.$$

***Claim*** **5.5.4.** *Under* (1), (2), (4) *with* $G_n \equiv G$, (68) – (70), (75) *and* (76), *the assumptions* (5) – (9) *are satisfied.*

**Proof.** By (41), (43), (49) and Theorems II.4.2.1 and V.1.3.1 of Hájek–Sĭdák (op. cit.),

(77) $$\lim_{x\to 0} \text{limsup } \max_i \int |f(\tau_i(y+x)) - f(y+x)|^r \, dG(y) = 0,$$

$$\lim_{x\to 0} \int |f(y+x) - f(y)|^r \, dG(y) = 0, \qquad r = 1, 2.$$

Now,

$$|\int \Sigma_i \|d_i\|^2 \{F_i(1-F_i) - F(1-F)\} \, dG|$$

$$\leq 2p\max_i \int |F(\tau_i y) - F(y)| \, dG(y) \leq 2p \max_i \int_1^{\tau_i} \int |y| \, f(sy) \, dG(y) \, ds$$

$$= o(1), \qquad \text{by (48) and (49) with } j = 1, r = 1.$$

Hence (69) implies (5). Next,

$$|\int \gamma_n(y+x) \, dG(y) - \Sigma_i \|d_i\|^2 \int f \, dG|$$

$$\leq \Sigma_i \|d_i\|^2 \tau_i \int \{|f(\tau_i(y+x)) - f(y+x)| + |f(y+x) - f(y)|\} dG(y) +$$

$$+ \max_i \sigma_i \, p \int f dG.$$

Therefore, in view of (48), (77) and (68) we obtain (6). Next, consider

$$\int \{f_i(y+x) - f_i(y)\}^2 \, dG(y)$$

$$\leq 4\tau_i^2 \int \Big\{[f(\tau_i(y+x)) - f(y+x)]^2 + [f(y+x) - f(y)]^2 +$$

$$+ [f(\tau_i y) - f(y)]^2\Big\} \, dG(y)$$

Therefore, (75) and (50) imply (39), and hence (7) and (9) by Claim 5.5.1. Note that (50) and (41b) imply (40). Finally,

$$\sum_{j=1}^{p} \int \|\nu_j - \sum_i d_{ij} \, c_i f\|^2 \, dG$$

$$\leq p^2 \max_i \int \{\tau_i f(\tau_i y) - f(y)\}^2 \, dG(y)$$

$$\leq 2p^2 \max_i \tau_i^2 \Big[\int \{f(\tau_i y) - f(y)\}^2 \, dG(y) + \int f^2 \, dG\Big] = o(1),$$

by (75), (70b), (77). Hence (70b) and the fact that $\sum_{j=1}^{p} \|\sum_i d_{ij} \, c_i\|^2 \leq p^2$ implies (8). □

Here, the assumption (10) is equivalent to having

(78) $$\sum_{j=1}^{p} \int [\sum_i d_{ij}\{F(\tau_i y) - F(y)\}]^2 \, dG(y) = O(1).$$

One sufficient condition for (78), besides requiring $F$ to have density $f$ satisfying

(79) $$\lim_{s\to 1} \int (yf(sy))^2 \, dG(y) = \int (yf(y))^2 \, dG(y) < \infty,$$

is to have

(80) $$\sum_{i=1}^{n} \sigma_i^2 = O(1).$$

One choice of $\{\sigma_i\}$ satisfying (80) is $\sigma_i^2 \equiv n^{-1/2}$ and the other choice is $\sigma_i^2 \propto \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$, $1 \le i \le n$.

Again, if $f$ satisfies **(F1)**, **(F3)** and $G$ is a finite measure then (68) (70), (76) and (79) are *a priori* satisfied. □

Now we shall give a set of sufficient conditions that will yield (5.4.A1) for the $Q$ of (5.2.13). Since $Q$ does not satisfy (5.3.21), the distribution of $Q$ under (1.1.1) is not idependent of $\boldsymbol{\beta}$. Therefore care has to be taken to exhibit this dependence clearly when formulating a theorem pertaining to $Q$. This of course complicates the presentation somewhat. As before with $\{H_{ni}\}$, $\{F_{ni}\}$ denoting the modeled and the actual d.f.'s of $\{e_{ni}\}$, define for $0 \le s \le 1$, $y \in \mathbb{R}$, $\mathbf{t} \in \mathbb{R}^p$,

(81) $$\bar{H}_n(s, y, \mathbf{t}) := n^{-1} \sum_{i=1}^{ns} H_{ni}(y - \mathbf{x}_{ni}'\mathbf{t}),$$

$$m_n(s, y) := n^{-1/2} \sum_{i=1}^{ns} \{F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta}) - H_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\},$$

$$\mathcal{M}_{1n}(s, y) := n^{-1/2} \sum_{i=1}^{ns} \{I(Y_{ni} \le y) - F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\},$$

$$d\alpha_n(s, y) := dL_n(s)\, dG_n(y).$$

Observe that

$$Q(\mathbf{t}) = \int [\mathcal{M}_{1n}(s, y) + m_n(s, y) - n^{1/2}\{\bar{H}_n(s, y, \mathbf{t}) - \bar{H}_n(s, y, \boldsymbol{\beta})\}]^2 \, d\alpha_n(s, y).$$

Note that the single integral is over the set $[0, 1] \times \mathbb{R}$.

Assume that $\{H_{ni}\}$ have densities $\{h_{ni}\}$ w.r.t $\lambda$ and set

$$(82)\qquad \bar{\mathbf{R}}_n(s, y) := n^{-1/2}\sum_{i=1}^{ns} \mathbf{x}_{ni}\, h_{ni}(y - \mathbf{x}_{ni}'\beta),$$

$$\overline{h_n^2}(y) := n^{-1}\sum_{i=1}^{n} h_{ni}^2(y - \mathbf{x}_{ni}'\beta), \qquad s\in[0, 1],\ y \in \mathbb{R},$$

$$\bar{\boldsymbol{\nu}}_n := \mathbf{A}\bar{\mathbf{R}}_n, \qquad \mathscr{B}_{1n} := \int \bar{\boldsymbol{\nu}}_n\, \bar{\boldsymbol{\nu}}_n'\, d\alpha_n.$$

Finally define, for $\mathbf{t} \in \mathbb{R}^p$,

$$(83)\qquad \hat{Q}(\mathbf{t}) := \int [M_{1n}(s, y) + m_n(s, y) + \mathbf{t}'\bar{\mathbf{R}}_n(s, y)]^2\, d\alpha_n(s, y).$$

**Theorem 5.5.7.** *Assume that* (1.1.1) *holds with the actual and the modeled d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$ *and* $\{H_{ni}, 1 \le i \le n\}$, *respectively. In addition, assume that* (1) *holds,* $\{H_{ni}, 1 \le i \le n\}$ *have densities* $\{h_{ni}, 1 \le i \le n\}$ *w.r.t.* $\lambda$, *and the following hold.*

$$(84)\qquad |\overline{h_n^2}|_n = O(1).$$

(85) $\forall\ \mathbf{v} \in \mathcal{N}(B), \forall\ \delta > 0,$

$$\limsup_n \max_i (2\delta\kappa_{ni})^{-1} \int_{a_{ni}}^{b_{ni}} \int h_{ni}^2(y - \mathbf{x}_{ni}'\beta + z)\, dG_n(y)\, dz$$

$$= \limsup_n \max_i \int h_{ni}^2(y - \mathbf{x}_{ni}'\beta)\, dG_n(y) < \infty,$$

*where* $a_{ni} = -\delta\kappa_{ni} - \mathbf{c}_{ni}'\mathbf{v}$, $b_{ni} = \delta\kappa_{ni} - \mathbf{c}_{ni}'\mathbf{v}$, $\kappa_{ni} = \|\mathbf{c}_{ni}\|$, $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$, $1 \le i \le n$.

(86) $\forall\ \mathbf{u}\in\mathcal{N}(B),$

$$\int \{n^{1/2}[\bar{H}_n(s, y, \beta + \mathbf{A}\mathbf{u}) - \bar{H}_n(s, y, \beta)] + \mathbf{u}'\bar{\boldsymbol{\nu}}_n\}^2\, d\alpha_n(s, y) = o(1).$$

$$(87)\qquad \int n^{-1}\sum_{i=1}^{n} F_{ni}(y - \mathbf{x}_{ni}'\beta)\,(1 - F_{ni}(y - \mathbf{x}_{ni}'\beta))\, dG_n(y) = O(1).$$

$$(88)\qquad \int m_n^2(s, y)\, d\alpha_n(s, y) = O(1).$$

*Then,* $\forall\ 0 < B < \infty$,

$$(89)\qquad E \sup_{\|\mathbf{u}\|\le B} |Q(\beta + \mathbf{A}\mathbf{u}) - \hat{Q}(\mathbf{A}\mathbf{u})| = o(1).$$

The details of the proof are similar to those of Theorem 5.5.1 and are left out as an exercise for interested readers.

An analogue of (51) for $\bar{\beta}$ will appear in the next section as Theorem 5.6a.3. Its asymptotic distribution in the case when the errors are correctly modeled to be i.i.d. will be also discussed there.

We shall end this section by stating analogues of some of the above results that will be useful when an unknown scale is also being estimated. To begin with, consider $K_D$ of (5.2.24). To simplify writing, let

$$K_D^o(s, u) := K_D((1+sn^{-1/2}), Au), \qquad s \in \mathbb{R},\ u \in \mathbb{R}^p. \tag{90}$$

Write $a_s := (1 + sn^{-1/2})$. Then from (5.2.24) and (90),

$$K_D^o(s, u) = \sum_{j=1}^p \int \{Y_j^o(ya_s, u) + \mu_j^o(ya_s, u) - \sum_i d_{ij}\, H_i(y)\}^2 dG_n(y) \tag{91}$$

where $H_i$ is the d.f. of $e_i$, $1 \le i \le n$, and where $\mu_j^o$, $Y_j^o$ are as in (9) and (13), respectively. Writing $\mu_j^o(y)$, $Y_j^o(y)$ etc. for $\mu_j^o(y, 0)$, $Y_j^o(y, 0)$ etc., we obtain

$$\begin{aligned} K_D^o(s, u) = \sum_{j=1}^p \int \{ & Y_j^o(ya_s, u) - Y_j^o(y) + \mu_j^o(ya_s) - \mu_j^o(y) - sy\nu_j^*(y) \\ & + Y_j^o(y) + u'\nu_j(y) + sy\nu_j^*(y) + m_j(y) \\ & + \mu_j^o(ya_s, u) - \mu_j^o(ya_s) - \mu'\nu_j(ya_s) \\ & + u'[\nu_j(ya_s) - \nu_j(y)]\}^2\, dG_n(y) \end{aligned} \tag{92}$$

where $\nu_j$ is as in (8) and

$$\nu_j^*(y) := n^{-1/2} \sum_i d_{nij}\, f_{ni}(y), \qquad 1 \le j \le p. \tag{93}$$

The representation (92) suggesting the following approximating candidate:

$$\hat{K}_D^o(s, u) := \sum_{j=1}^p \int \{Y_j^o + u'\nu_j + sy\nu_j^* + m_j\}^2\, dG_n. \tag{94}$$

We now state

**Lemma 5.5.5.** *With* $\gamma_n$ *as in* (6), *assume that* $\forall\ |s| \le b,\ 0 < b < \infty$,

$$\begin{aligned} & \lim_{x \to 0} \limsup_n \int \gamma_n((1+sn^{-1/2})y+x)dG_n(y) \\ & \qquad = \limsup_n \int \gamma_n(y)dG_n(y) < \infty, \end{aligned} \tag{95}$$

*and*

(96) $$\lim_{z\to 0} \limsup_n \int |y|\, \gamma_n(y+zy)\, dG_n(y) = \limsup_n \int |y|\, \gamma_n(y)\, dG_n(y) < \infty.$$

*Moreover, assume that* $\forall\ (s, \mathbf{v}) \in [-b, b] \times \mathcal{N}(B) =: \mathcal{N}_1$, *and* $\forall\ \delta > 0$

(97) $$\limsup_n \sum_{j=1}^{p} \int \Big[\sum_{i=1}^{n} d_{nij}^{\pm}\{F_{ni}(ya_s + c_{ni}'\mathbf{v} + \delta(n^{-1/2}|y| + \kappa_{ni})) - F_{ni}(ya_s + c_{ni}'\mathbf{v} - \delta(n^{-1/2}|y| + \kappa_{ni}))\}\Big]^2 dG_n(y) \le k\delta^2,$$

*for some* k *not depending on* $(s, \mathbf{v})$ *and* $\delta$.
*Then,* $\forall\ 0 < b, B < \infty$,

(98) $$E \sup \sum_{j=1}^{p} \int \{Y_j^o((1+sn^{-1/2})y, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n = o(1)$$

*where the supremum is taken over* $(s, \mathbf{u}) \in \mathcal{N}_1$.

**Proof.** For each $(s, \mathbf{u}) \in \mathcal{N}_1$, with $a_s = 1 + sn^{-1/2}$,

$$E \sum_{j=1}^{p} \int \{Y_j^o(ya_s, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n(y)$$
$$\le \int_{-B_n}^{B_n} \int \gamma_n(ya_s+z)\, dG_n(y)dz + \int_{-b_n}^{b_n} \int |y|\, \gamma_n(y+zy)\, dG_n(y)dz$$

where $B_n = B \max_i \|\kappa_i\|$, $b_n = bn^{-1/2}$. Therefore, from (95) and (96), for every $(s, \mathbf{u}) \in \mathcal{N}_1$,

$$E \sum_{j=1}^{p} \int \{Y_j^o(ya_s, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n(y) = o(1).$$

Now proceed as in the proof of (16), using the monotonicity of $V_{jd}(a, \mathbf{t})$, $\mu_j^o(a, \mathbf{u})$ and the compactness of $\mathcal{N}_1$ to conclude (98). Use (97) in place of (7). The details are left out as an exercise. □

The proof of the following lemma is quite similar to that of (30).

**Lemma 5.5.6.** *Let* $G_n^*(y) = G_n(y/a_\tau)$. *Assume that for each fixed* $(\tau, \mathbf{u}) \in \mathcal{N}_1$, (8) *and* (9) *hold with* $G_n$ *replaced by* $G_n^*$. *Moreover, assume the following:*

(99) $$\sum_{j=1}^{p} \int (y\nu_j^*(y))^2 \, dG_n(y) = O(1).$$

(100) $$\sum_{j=1}^{p} \int \{\mu_j^o(ya_s) - \mu_j^o(y) - \tau y\nu_j^*(y)\}^2 \, dG_n(y) = o(1), \quad \forall \ |s| \le b.$$

*Then,* $\forall\ 0 < b,\ B < \infty$,

(101) $$\sup \sum_{j=1}^{p} \int \{\mu_j^o(ya_s, \mathbf{u}) - \mu_j^o(a_\tau y) - \mathbf{u}'\nu_j(a_\tau y)\}^2 \, dG_n(y) = o(1),$$

*and*

(102) $$\sup \sum_{j=1}^{p} \int \{\mu_j^o(ya_s) - \mu_j^o(y) - \tau y\nu_j^*(y)\}^2 \, dG_n(y) = o(1).$$

*where the supremum in* (101), (102) *is taken over* $(s, \mathbf{u}) \in \mathcal{N}_1$, $|s| \le b$, *respectively.*

**Theorem 5.5.8.** *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Assume* (1) – (5), (8), (10), (95) – (97) *and the conditions of Lemma* 5.5.6 *hold. Moreover assume that for each* $|s| \le b$

(103) $$\sum_{j=1}^{p} \int \|\nu_j(ya_s) - \nu_j(y)\|^2 \, dG_n(y) = o(1).$$

*Then,* $\forall\ 0 < b,\ B < \infty$,

(104) $$E \sup |K_D^o(\tau, \mathbf{u}) - \hat{K}_D^o(\tau, \mathbf{u}) = o(1).$$

*where the supremum is taken over* $(s, \mathbf{u}) \in \mathcal{N}_1$.

The proof of this theorem is quite similar to that of Theorem 5.5.1. □

## 5.6. ASYMPTOTIC DISTRIBUTIONS, EFFICIENCES AND ROBUSTNESS

### 5.6a. Asymptotic Distributions and Efficiences

To begin with consider the *Case* 5.5.1 *and the class of estimators* $\{\hat{\beta}_D\}$. Recall that in this case the errors $\{e_{ni}\}$ of (1.1.1) are correctly modeled to be i.i.d. $F$, i.e., $H_{ni} \equiv F \equiv F_{ni}$. We shall also take $G_n \equiv G$, $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$. Assume that (5.5.68) – (5.5.70) hold. The various quantities appearing in (5.5.37) and Theorem 5.5.3 now take the following simpler forms.

$$(1) \qquad \Gamma_n(y) = A\,X'D\,f(y),\ y \in \mathbb{R}, \qquad \mathscr{B}_n = AX'D\,D'XA \int f\,dG,$$

$$\mathscr{T}_n = -AX'D \int Y^o_D\, f\,dG.$$

Note that $\mathscr{B}_n^{-1}$ will exist if and only if the rank of $D$ is $p$. Note also that

$$(2) \qquad \mathscr{B}_n^{-1}\mathscr{T}_n = -(D'XA)^{-1} \int Y^o_D\, f dG \,/\, (\int f^2 dG)^{-1}$$
$$= (D'XA \int f^2 dG)^{-1}\, \Sigma_i\, d_i\, [\psi(e_i) - E\psi(e_i)],$$

where $\psi(y) = \int_{-\infty}^{y} f dG, \quad y \in \mathbb{R}$.

Because $G_n \equiv G \in \mathcal{DI}(\mathbb{R})$, there always exists a $g \in L_r^2(G)$ such that $g > 0$, and $0 < \int g^2 dG < \infty$. Take $g_n \equiv g$ in (5.5.11). Then the condition (5.5.11) translates to assuming that

$$(3) \qquad \liminf_n \inf_{\|\theta\|=1} |\theta' D'XA\theta| \geq \alpha \qquad \text{for some } \alpha > 0.$$

Condition (5.5.12) implies that $\theta' D'XA\theta \geq 0$ or $\theta' D'XA\,\theta \leq 0$, $\forall\ \|\theta\| = 1$ and $\forall\ n \geq 1$. It need not imply (3). The above discussion together with the L-F Cramer-Wold Theorem leads to

**Corollary 5.6a.1.** *Assume that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F, F *known. In addition, assume that* (5.5.1), (5.5.2), (5.5.12), (5.5.68) – (5.5.70), (3) *and* (4) *hold, where*

$$(4) \qquad (D'XA)^{-1} \textit{ exists for all } n \geq p.$$

*Then,*

$$(5) \qquad A^{-1}(\hat\beta_D - \beta) = (D'XA \int f^2 dG)^{-1} \sum_{i=1}^{n} d_{ni}\, [\psi(e_{ni}) - E\psi(e_{ni})] + o_p(1).$$

*If, in addition, we assume*

$$(6) \qquad \max_{1 \leq i \leq n} \|d_{ni}\|^2 = o(1),$$

*then*

$$(7) \qquad \Sigma_D^{-1} A^{-1}(\hat\beta_D - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p})$$

*where*

$$\Sigma_D := (D'XA)^{-1} D'D(AX'D)^{-1}, \qquad \tau^2 = \mathrm{Var}\, \psi(e_1)/(\int f^2 dG)^2. \qquad \square$$

For any two square matrices $L_1$ and $L_2$ of the same order, by $L_1 \geq L_2$ we mean that $L_1 - L_2$ is non–negative definite. Let $L$ and $J$ be two $p \times n$ matrices such that $(LL')^{-1}$ exists. The C–S inequality for matrices states that

(8) $$JJ' \geq JL'(LL')^{-1}LJ' \text{ with equality if and only if } J \propto L.$$

Now note that if $D = XA$ then $\Sigma_D = I_{p \times p}$. In general, upon choosing $J = D'$, $L = AX'$ in (8), we obtain

$$D'D \geq D'XA \cdot AX'D \quad \text{or} \quad \Sigma_D \geq I_{p \times p}$$

with equality if and only if $D \propto XA$. From these observations we deduce

**Theorem 5.6a.1.** (***Optimality of*** $\hat{\beta}_X$). *Suppose that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F. *In addition, assume that* (5.5.1), (5.5.4) *with* $G_n \equiv G$, (5.5.68) – (5.5.70) *hold. Then, among the class of estimators* $\{\hat{\beta}_D$; D *satisfying* (5.5.2), (5.5.12), (3), (4) *and* (5)$\}$, *the estimator that minimizes the asymptotic variance of* $b'A^{-1}(\hat{\beta}_D - \beta)$, *for every* $b \in \mathbb{R}^p$, *is* $\hat{\beta}_X$ *– the* $\hat{\beta}_D$ *with* $D = XA$. □

Observe that under (5.5.1), $D = XA$ *a priori* satisfies (5.5.2), (3), (4) and (6). Consequently we obtain

**Corollary 5.6a.2.** (***Asymptotic normality of*** $\hat{\beta}_X$.) *Assume that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F. *In addition, assume that* (5.5.1) and (5.5.68) – (5.5.70) *hold. Then,*

$$A^{-1}(\hat{\beta}_X - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p \times p}).$$ □

**Remark 5.6a.1.** Write $\hat{\beta}_D(G)$ for $\hat{\beta}_D$ to emphasize the dependence on G. The above theorem proves the optimality of $\hat{\beta}_X(G)$ among a class of estimators $\{\hat{\beta}_D(G)$, as D varies$\}$. To obtain an asymptotically efficient estimator at a given F among the class of estimators $\{\hat{\beta}_X(G)$, G varies$\}$ one must have F and G satisfy the following relation. Assume that F satisfies (3.2.a) of Theorem 3.2.3 and all of the derivatives that occur below make sense and that (5.5.68) hold. Then, a G that will give asymptotically efficient $\hat{\beta}_X(G)$ must satisfy the relation

$$-f\,dG = (1/I(f))\cdot d(\dot{f}/f), \qquad I(f) := \int (\dot{f}/f)^2 dF.$$

From this it follows that the m.d. estimators $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$, for $G$ satisfying the relations $dG(y) = (2/3)dy$ and $dG(y) = 4d\delta_0(y)$, are asymptotically efficient at logistic and double exponential error d.f.'s, respectively.

For $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$ to be asymptotically efficient at $N(0, 1)$ errors, $G$ would have to satisfy $f(y)dG(y) = dy$. But such a $G$ does not satisfy (5.5.58). Consequently, under the current art of affairs, one can not estimate $\boldsymbol{\beta}$ asymptotically efficiently at the $N(0, 1)$ error d.f. by using a $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$. This naturally leaves one open problem, v.i.z., *Is the conclusion of Corollary* 5.6a.2 *true without requiring* $\int f dG < \infty$, $0 < \int f^2 dG < \infty$? □

Observe that Theorem 5.6a.1 does not include the estimator $\hat{\beta}_1$ – the $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ when $\mathbf{D} = n^{1/2}[1, 0, \ldots, 0]_{n\times p}$ i.e., the m.d. estimator defined at (5.2.4), (5.2.5) after $H_{ni}$ is replaced by $F$ in there. The main reason for this being that the given $\mathbf{D}$ does not satisfy (4). However, Theorem 5.5.3 is general enough to cover this case also. Upon specializing that theorem and applying (5.5.49) one obtains the following

**Theorem 5.6a.2.** *Assume that* (1.1.1) *holds with the errors correclty modeled to be i.i.d.* F. *In addition, assume that* (5.5.1), (5.5.68) – (5.5.70) *and the following hold.*

(10) *Either*

$$n^{-1/2}\theta_1\mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \ge 0 \;\; \textit{for all} \;\; 1 \le i \le n, \;\; \textit{all} \;\; \|\boldsymbol{\theta}\| = 1,$$

*or*

$$n^{-1/2}\theta_1\mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \le 0 \;\; \textit{for all} \;\; 1 \le i \le n, \;\; \textit{all} \;\; \|\boldsymbol{\theta}\| = 1.$$

$$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |n^{1/2}\theta_1\bar{\mathbf{x}}_n'\mathbf{A}\boldsymbol{\theta}| \ge \alpha > 0, \tag{11}$$

*where* $\bar{\mathbf{x}}_n$ *is as in* (4.2a.11) *and* $\theta_1$ *is the first coordinate of* $\boldsymbol{\theta}$. *Then*

$$n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}\cdot\mathbf{A}^{-1}(\hat{\beta}_1 - \boldsymbol{\beta}) = Z_n \Big/ \int f^2 dG + o_p(1), \tag{12}$$

*where*

$$Z_n = n^{-1/2}\,\Sigma_i\,\{\psi(e_{ni}) - E\psi(e_{ni})\}, \quad \textit{with } \psi \textit{ as in } (2).$$

*Consequently,* $n^{1/2}\bar{\mathbf{x}}_n'(\hat{\beta}_1 - \boldsymbol{\beta})$ *is asymptotically a* $N(0, \tau^2)$ *r.v.* □

Next, we focus on the class of estimators $\{\beta^+_{\mathbf{D}}\}$ and the *case of i.i.d. symmetric errors.* An analogue of Corollary 5.6a.1 is obtained with the help of Theorem 5.5.4 instead of Theorem 5.5.3 and is given in Corollary 5.6a.3. The details of its proof are similar to those of Corollary 5.6a.1.

**Corollary 5.6a.3.** *Assume that* (1.1.1) *holds with the errors correctly modeled to be i.i.d. symmetric around* 0. *In addition, assume that* (5.3.8), (5.5.1), (5.5.2), (5.5.4) *with* $G_n \equiv G$, (5.5.68), (5.5.70), (3), (4) *and* (13) *hold, where*

$$(13) \qquad \int_0^\infty (1-F)\, dG < \infty$$

*Then,*

$$(14) \qquad \mathbf{A}^{-1}(\beta^+_{\mathbf{D}} - \beta) = -\{2\mathbf{A}\mathbf{X}'\mathbf{D}\int f^2 dG\}^{-1}.\int \mathbf{W}^+(y)\, f^+(y)dG(y) + o_p(1),$$

*where* $f^+(y) := f(y) + f(-y)$ *and* $\mathbf{W}^+(y)$ *is* $\mathbf{W}^+(y, 0)$ of (5.5.32). *If, in addition,* (6) *holds, then*

$$(15) \qquad \Sigma_{\mathbf{D}}^{-1}\mathbf{A}^{-1}(\beta^+_{\mathbf{D}} - \beta) \underset{d}{\longrightarrow} N(0, \tau^2 I_{p\times p}). \qquad \square$$

Consequently, an analogue of Theorem 5.6a.1 holds for $\beta^+_{\mathbf{X}}$ also and Remark 5.6a.1 applies equally to the class of estimators $\{\beta^+_{\mathbf{X}}(G),\ G \text{ varies}\}$, assuming that the errors are symmetric around 0. We leave it to interested readers to state and prove an analogue of Theorem 5.6a.2 for $\beta^+_1$.

Now consider the class of estimators $\{\beta^*_{\mathbf{D}}\}$ of (5.2.23). Recall the notation in (5.5.61) and Theorem 5.5.6. The distributions of these estimators will be discussed when the errors in (1.1.1) are correctly modeled to be i.i.d. F, F an arbitrary d.f. and when $L_n \equiv L$. In this case various entities of Theorem 5.5.6 acquire the following forms.

$$\mu_{\mathbf{D}} \equiv 0; \qquad \ell_{ni}(s) \equiv 1; \qquad \mathbf{D}(s) \equiv \mathbf{D}, \text{ under (5.2.21)};$$
$$\Gamma^*_n(s) \equiv A_1 X_c' \mathbf{D}\, q(s), \qquad q = f(F^{-1});$$
$$\mathcal{T}^*_n = -A_1 X_c' \mathbf{D} \int \mathbf{Y}_{\mathbf{D}}\, q dL = A_1 X_c' \mathbf{D} \sum_{i=1}^n d_{ni}\, \varphi_0(F(e_{ni}));$$
$$\mathcal{B}^*_n = (A_1 X_c' \mathbf{D}\, \mathbf{D}' X_c A_1) \int q^2 dL,$$

where $\mathbf{X}_c$ and $A_1$ are defined at (4.2a.11) and where

$$\varphi_0(u) := \int_0^u q(s)dL(s), \qquad 0 \le u \le 1.$$

Arguing as for Corollary 5.6a.1, one obtains the following

**Corollary 5.6a.4.** *Assume that* (1.1.1) *holds with the errors correctly modeled to be i.i.d.* F *and that* L *is a d.f.. In addition, assume that* (**F1**), (**NX**$_c$), (5.2.21), (5.5.2), *and the following hold.*

(16) $$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |\boldsymbol{\theta}'\mathbf{D}'\mathbf{X}_c\mathbf{A}_1\boldsymbol{\theta}| \geq \alpha > 0$$

(17) Either

$$\boldsymbol{\theta}'\mathbf{d}_{ni}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\mathbf{A}_1\boldsymbol{\theta} \geq 0, \quad \forall\ 1 \leq i \leq n, \quad \forall\ \|\boldsymbol{\theta}\| = 1,$$

or

$$\boldsymbol{\theta}'\mathbf{d}_{ni}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\mathbf{A}_1\boldsymbol{\theta} \leq 0, \quad \forall\ 1 \leq i \leq n, \quad \forall\ \|\boldsymbol{\theta}\| = 1.$$

(18) $$(\mathbf{D}'\mathbf{X}_c\mathbf{A}_1)^{-1} \text{ exists for all } n \geq p.$$

*Then,*

(19) $$\mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{D}}^* - \boldsymbol{\beta}) = (\mathbf{D}'\mathbf{X}_c\mathbf{A}_1 \int_0^1 q^2 dL)^{-1} \sum_{n=1}^{n} \mathbf{d}_{ni}\, \varphi(F(e_{ni})) + o_p(1).$$

*If, in addition,* (6) *holds, then*

(20) $$(\boldsymbol{\Sigma}_{\mathbf{D}}^*)^{-1}\mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{D}}^* - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N(\mathbf{0}, \sigma_0^2 \mathbf{I}_{p\times p})$$

*where* $\boldsymbol{\Sigma}_{\mathbf{D}}^* = (\mathbf{D}'\mathbf{X}_c\mathbf{A}_1)^{-1}\mathbf{D}'\mathbf{D}(\mathbf{A}_1\mathbf{X}_c'\mathbf{D})^{-1}$, $\quad \sigma_0^2 = \mathrm{Var}\, \varphi(F(e_1))/(\int_0^1 q^2 dL)^2$.
*Consequently,*

(21) $$\mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{X}_c}^* - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N(\mathbf{0}, \sigma_0^2 \mathbf{I}_{pxp})$$

*and* $\{\boldsymbol{\beta}_{\mathbf{X}_c}^*\}$ *is asymptotically efficient among all* $\{\boldsymbol{\beta}_{\mathbf{D}}^*$, **D** *satisfying above conditions*$\}$. □

Consider the case when $L(s) \equiv s$. Then

$$\sigma_0^2 = (\int f^3(x)dx)^{-2} \int\int [F(x \wedge y) - F(x)F(y)] f^2(x) f^2(y)\, dxdy.$$

It is interesting to make a numerical comparison of this variance with that of some other well celebrated estimators. Let $\sigma_w^2$, $\sigma_{lad}^2$, $\sigma_{ls}^2$ and $\sigma_{ns}^2$ denote the respective asymptotic variances of the Wilcoxon rank, the least absolute deviation, the least square and the normal scores estimators of $\boldsymbol{\beta}$. Recall, either from Chapter 4 or from Jaeckel (1972) that

$$\sigma_w^2 = (1/12)\cdot\{\int f^2(x)\,dx\}^{-2}; \quad \sigma_{lad}^2 = (2\,f(0))^{-2}; \quad \sigma_{ls}^2 = \sigma^2;$$

$$\sigma_{ns}^2 = \{\int f^2(x)/\varphi(\Phi^{-1}(F)))\,dx\}^{-2};$$

where $\sigma^2$ is the error variance. Using these we obtain the following table.

**Table I**

| F \ $\sigma_\cdot^2$ | $\sigma_0^2$ | $\sigma_w^2$ | $\sigma_{lad}^2$ | $\sigma_{ns}^2$ | $\sigma^2$ |
|---|---|---|---|---|---|
| Double Exp. | 1.2 | 1.333 | 1 | $\pi/2$ | 2 |
| Logistic | 3.0357 | 3 | 4 | $\pi$ | $\pi^2/3$ |
| Normal | 1.0946 | $\pi/3$ | $\pi/2$ | 1 | 1 |

It thus follows that the m.d. estimator $\beta^*_{X_c}(L)$, with $L(s) \equiv s$, is superior to the Wilcoxon rank estimator and the l.a.d. estimator at double exponential and logistic errors, respectively. At normal errors, it has smaller variance than the l.a.d. estimator and compares favorably with the optimal estimator. The same is true for the m.d. estimator $\hat\beta_X(F)$.

Next, we shall discuss $\bar\beta$. In the following theorem the framework is the same as in Theorem 5.5.7. Also see (5.5.82) for the definitions of $\bar\nu_n$, $\mathcal{B}_{ln}$ etc.

**Theorem 5.6a.3.** *In addition to the assumptions of Theorem* 5.5.7 *assume that*

(22) $$\liminf_n \inf_{\|\theta\|=1} |\int \bar\nu_n'\, d\alpha_n\theta| \geq \alpha, \qquad \textit{for some } \alpha > 0.$$

*Moreover, assume that* (10) *holds and that*

(23) $\mathcal{B}_{ln}^{-1}$ *exists for all* $n \geq p$.

*Then,*

(24) $$A^{-1}(\bar\beta - \beta) = -\mathcal{B}_{ln}^{-1}\int\int \bar\nu_n(s, y)\{\mathcal{M}_{ln}(s, y) + m_n(s, y)\}\, d\alpha_n(s, y) + o_p(1).$$

**Proof.** The proof of (23) is similar to that of (5.5.51), hence no details are given. □

**Corollary 5.6a.5.** *Suppose that the conditions of Theorem* 5.6a.3 *are satisfied by* $F_{ni} \equiv F \equiv H_{ni}$, $G_n \equiv G$, $L_n \equiv L$, *where* F *is supposed to have* *continuous density* f. *Let*

$$(25)\quad C = \int\int\int_0^1\int_0^1 [\{A\, n^{-1} \sum_{i=1}^{ns}\sum_{j=1}^{nt} \mathbf{x}_i\mathbf{x}_j' f_i(y) f_j(y)\, A\}\,(s\wedge t)\cdot$$
$$\cdot\{F(y\wedge z) - F(y)F(z)\}]\, d\alpha(s, y)\, d\alpha(t, z),$$

*where* $f_i(y) = f(y - \mathbf{x}_i'\beta)$, *and* $d\alpha(s, y) = dL(s)dG(y)$. *Then the asymptotic distribution of* $A^{-1}(\bar\beta - \beta)$ is $N(0, \Sigma_0(\beta))$ *where* $\Sigma_0(\beta) = \mathscr{B}_{1n}^{-1} C \mathscr{B}_{1n}^{-1}$. □

Because of the dependence of $\Sigma_0$ on $\beta$, no clear cut comparison between $\bar\beta$ and $\hat\beta_{\mathbf{X}}$ in terms of their asymptotic covariance matrices seems to be feasible. However, some comparison at a given $\beta$ can be made. To demonstrate this, consider the case when $L(s) = s$, $p = 1$ and $\beta_1 = 0$. Write $x_i$ for $x_{i1}$ etc.

Note that here, with $\tau_x^2 = \sum_{i=1}^n x_i^2$,

$$\mathscr{B}_{1n} = \tau_x^{-2} \int_0^1 n^{-1} \sum_{i=1}^{ns} x_i \sum_{j=1}^{ns} x_j\, ds \cdot \int f^2\, dG,$$

$$C = \tau_x^{-2} \int_0^1\int_0^1 n^{-1} \sum_{i=1}^{ns} x_i \sum_{j=1}^{nt} x_j\,(s\wedge t)\, ds dt\cdot$$
$$\cdot\int\int [F(y\wedge z) - F(y)F(z)]\, d\psi(y)\, d\psi(z).$$

Consequently

$$\Sigma_0(0) = \frac{\tau_x^{-2} \int_0^1\int_0^1 (s\wedge t) n^{-1} \sum_{i=1}^{ns} x_i \sum_{j=1}^{nt} x_j\, ds dt}{\left(\tau_x^{-2} \int_0^1 n^{-1} \sum_{i=1}^{ns} x_i \sum_{j=1}^{ns} x_j\, ds\right)^2} . \tau^2 = r_n\cdot\tau^2, \qquad \text{say.}$$

Recall that $\tau^2$ is the asymptotic variance of $\tau_x(\hat\beta_x - \beta)$. Direct integration shows that in the cases $x_i \equiv 1$ and $x_i \equiv i$, $r_n \longrightarrow 18/15$ and $50/21$, respectively. Thus, in the cases of the one sample location model and the first degree polynomial through the origin, in terms of the asymptotic variance, $\hat\beta_x$ dominates $\bar\beta$ with $L(s) = s$ at $\beta = 0$. □

### 5.6b. Robustness

In a linear regression setup an estimator needs to be robust against departures in the assumed design variables and the error distributions. As seen in Section 5.6a, one purpose of having general weights $\mathbf{D}$ in $\hat{\beta}_{\mathbf{D}}$ was to prove that $\hat{\beta}_{\mathbf{X}}$ is asymptotically efficient among a certain class of m.d. estimators $\{\hat{\beta}_{\mathbf{D}}, \mathbf{D}$ varies$\}$. Another purpose is to robustify these estimators against the extremes in the design by choosing $\mathbf{D}$ to be a bounded function of $\mathbf{X}$ that satisfies all other conditions of Theorem 5.6a.1. Then the corresponding $\hat{\beta}_{\mathbf{D}}$ would be asymptotically normal and robust against the extremes in the design, but not as efficient as $\hat{\beta}_{\mathbf{X}}$. This gives another example of the phenomenon that compromises efficency in return for robustness. A similar remark applies to $\{\beta^{+}_{\mathbf{D}}\}$ and $\{\beta^{*}_{\mathbf{D}}\}$.

We shall now focus on the *qualitative robustness* (see Definition 4.4.1) of $\hat{\beta}_{\mathbf{X}}$ and $\beta^{+}_{\mathbf{X}}$. For simplicity, we shall write $\hat{\beta}$, $\beta^{+}$, for $\hat{\beta}_{\mathbf{X}}$, $\beta^{+}_{\mathbf{X}}$ in the rest of the section. To begin with consider $\hat{\beta}$. Recall Theorem 5.5.3 and the notation of (5.5.37). We need to apply these to the case when the errors in (1.1.1) are modeled to be i.i.d. F, but their actual d.f.'s are $\{F_{ni}\}$, $\mathbf{D} = \mathbf{XA}$ and $G_n = G$. Then various quantities in (5.5.37) acquire the following form.

$$(1) \qquad \Gamma_n(y) = \mathbf{AX}'\Lambda^{*}(y)\mathbf{XA}, \qquad\qquad \mathcal{B}_n = \mathbf{AX}'\int \Lambda^{*}\Pi\Lambda^{*}\, dG\, \mathbf{XA},$$

$$\mathcal{T}_n = \int \Gamma_n(y)\mathbf{AX}'[\boldsymbol{\alpha}_n(y) + \boldsymbol{\Delta}_n(y)]\, dG(y) = \mathbf{Z}_n + \mathbf{b}_n, \qquad \text{say,}$$

where

$$(2) \qquad \Pi := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'; \qquad\qquad \mathbf{b}_n := \int \Gamma_n(y)\mathbf{AX}'\boldsymbol{\Delta}_n(y)\, dG(y);$$

$$\alpha_{ni}(y) := I(e_{ni} \le y) - F_{ni}(y),$$

$$\Delta_{ni}(y) := F_{ni}(y) - F(y), \qquad\qquad 1 \le i \le n, \ y \in \mathbb{R};$$

$$\boldsymbol{\alpha}_n' := (\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nn}), \qquad \boldsymbol{\Delta}_n' := (\Delta_{n1}, \Delta_{n2}, \ldots, \Delta_{nn}).$$

The assumption (5.2.1) ensures that the design matrix $\mathbf{X}$ is of the full rank p. This in turn implies the existence of $\mathcal{B}_n^{-1}$ and the satisfaction of (5.2.2), (5.2.12) in the present case. Moreover, because $G_n \equiv G$, (5.2.11) now becomes

(3) $$\liminf_n \inf_{\|\theta\|=1} k_n(\theta) \geq \gamma, \qquad \text{for some } \gamma > 0,$$

where

$$k_n(\theta) := \theta' AX' \int \Lambda^* \, g dG \, XA\theta, \qquad \|\theta\| = 1,$$

and where $g$ is a function from $\mathbb{R}$ to $[0, \infty]$, $0 < \int g^r \, dG < \infty$, $r = 1, 2$. Because $G$ is a $\sigma$–finite measure, such a $g$ always exists.

Upon specializing Theorem 5.5.3 to the present case, we readily obtain

**Corollary 5.6b.1.** *Assume that* in (1.1.1) *the actual and modeled d.f.'s of the errors* $\{e_{ni}, 1 \leq i \leq n\}$ *are* $\{F_{ni}, 1 \leq i \leq n\}$ *and* $F$, *respectively. In addition, assume that* (5.5.1), (5.5.3) – (5.5.10) *with* $D = XA$, $H_{ni} \equiv F$, $G_n \equiv G$, *and* (3) *hold. Then*

(4) $$A^{-1}(\hat{\beta} - \beta) = -\mathscr{B}_n^{-1}\{Z_n + b_n\} + o_p(1). \qquad \square$$

Observe that $\mathscr{B}_n^{-1} b_n$ measures the amount of the asymptotic bias in the estimator $\hat{\beta}$ when $F_{ni} \neq F$. Our goal here is to obtain the asymptotic distribution of $A^{-1}(\hat{\beta} - \beta)$ when $\{F_{ni}\}$ converge to $F$ in a certain sense. The achievement of this goal is facilitated by the following lemma. Recall that for any square matirx $L$, $\|L\|_\infty = \sup\{\|t'L\|;\ \|t\| \leq 1\}$. Also recall the fact that

(5) $$\|L\|_\infty \leq \{\text{tr.} LL'\}^{1/2},$$

where tr. denotes the trace operator.

**Lemma 5.6b.1.** *Let* $F$ *and* $G$ *satisfy* (5.5.68). *Assume that* (5.5.5) *and* (5.5.10) *are satisfied by* $G_n \equiv G$, $\{F_{ni}\}$, $H_{ni} \equiv F$ *and* $D = XA$. *Moreover assume that* (5.5.3) *holds and that*

(6) $$\rho_n := \int (\Sigma_i \|x_{ni}'A\|^2 \, |f_{ni} - f|)^2 dG = o(1).$$

*Then with* $I = I_{p \times p}$,

(i) $$\| \mathscr{B}_n - I \int f^2 dG \|_\infty = o(1).$$

(ii) $$\| \mathscr{B}_n^{-1} - I(\int f^2 dG)^{-1} \|_\infty = o(1).$$

(iii) $$|\text{tr.}\ \mathscr{B}_n - p \int f^2 dG| = o(1).$$

(iv) $$\Big|\sum_{j=1}^{p}\int\|\nu_j\|^2dG - p\int f^2dG\Big| = o(1).$$

(v) $$\|b_n - \int AX'\Delta_n(y)f(y)dG(y)\| = o(1).$$

(vi) $$\|Z_n - \int AX'\alpha_n(y)f(y)dG(y)\| = o_p(1).$$

(vii) $$\sup_{\|\theta\|=1}\Big|k_n(\theta) - \int f\,g dG\Big| = o(1).$$

**Remark 5.6b.1.** Note that the condition (5.5.10) with $D = XA$, $G_n \equiv G$ now becomes

(7) $$\int\|AX'\Delta_n\|^2dG = O(1).$$

**Proof.** To begin with, because $AX'XA \equiv I$, we obtain the relation

$$\begin{aligned}\Gamma_n(y)\Gamma_n'(y) - f^2(y)I &= AX'[\Lambda^*(y) - f(y)I]XA\cdot AX'[\Lambda^*(y) - f(y)I]XA \\ &= AX'\mathcal{C}(y)XA\cdot AX'\mathcal{C}(y)XA \\ &= \mathcal{D}(y)\mathcal{D}'(y), \qquad y \in \mathbb{R},\end{aligned}$$

where $\mathcal{C}(y) := \Lambda^*(y) - I\,f(y)$, $\mathcal{D}(y) := AX'\mathcal{C}(y)XA$, $y \in \mathbb{R}$. Therefore,

(8) $$\|\mathcal{B}_n - I\int f^2dG\|_\infty \le \sup_{\|t\|\le 1}\int\|t'\mathcal{D}(y)\mathcal{D}'(y)\|\,dG(y) \le \int\{tr.LL'\}^{1/2}dG$$

where $L = \mathcal{D}\mathcal{D}'$. Note that, by the C–S inequality,

(9) $$tr.\,LL' = tr.\,\mathcal{D}\mathcal{D}'\mathcal{D}'\mathcal{D} \le \{tr.\mathcal{D}\mathcal{D}'\}^2.$$

Let $\delta_i = f_i - f$, $1 \le i \le n$. Then

(10) $$\begin{aligned}|tr.\mathcal{D}\mathcal{D}'| &= |tr.\,\Sigma_i\Sigma_j\,Ax_i\,x_i'A\cdot Ax_jx_j'A\cdot\delta_i\delta_j| \\ &= |\Sigma_i\Sigma_j\,\delta_i\delta_j\,(x_j'AAx_i)^2| \\ &\le \Sigma_i\Sigma_j\,|\delta_i\delta_j|\cdot\|x_i'A\|^2\cdot\|x_j'A\|^2 \\ &= (\Sigma_i\|Ax_i\|^2|\delta_i|)^2 = \rho_n.\end{aligned}$$

Consequently, from (8) – (10),

$$\|\mathscr{B}_n - \mathbf{I}\int f^2 dG\|_\infty \le \int (\Sigma_i \|\mathbf{A}\mathbf{x}_i\|^2 |f_i - f|)^2 dG = o(1), \quad \text{by (6).} \tag{11}$$

This proves (i) while (ii) follows from (i) by using the determinant and cofactor formula for the inverses.

Next, (iii) follows from (6) and the fact that

$$|\mathrm{tr}.\mathscr{B}_n - p\int f^2 dG| = |\int \mathrm{tr}.\mathscr{D}\mathscr{D}'\ dG| \le \rho_n, \qquad \text{by (10).} \tag{12}$$

To prove (iv), note that with $\mathbf{D} = \mathbf{XA}$,

$$\sum_{j=1}^{p}\int \|\nu_j\|^2 dG = \sum_{i=1}^{n}\sum_{k=1}^{n}\int \mathbf{x}_i'\mathbf{A}\mathbf{A}\mathbf{x}_k\ \mathbf{x}_k'\mathbf{A}\mathbf{A}\mathbf{x}_i\ f_i(y)f_k(y)\ dG(y).$$

Note that the r.h.s. is $p\int f^2 dG$ in the case $f_i \equiv f$. Thus

$$|\sum_{j=1}^{p}\int \|\nu_j\|^2\ dG - p\int f^2 dG| = |\int \mathrm{tr}.\mathscr{D}\mathscr{D}' dG| \le \rho_n. \tag{13}$$

This and (6) proves (iv).

Similarly, with $\boldsymbol{d}_j'(y)$ denoting the $j^{\text{th}}$ row of $\mathscr{D}(y)$, $1 \le j \le p$,

$$\begin{aligned} \|\mathbf{b}_n - \int \mathbf{A}\mathbf{X}'\Delta_n\, f dG\|^2 &= \|\int \mathscr{D}\mathbf{A}\mathbf{X}'\Delta\ dG\|^2 \\ &= \sum_{j=1}^{p}\{\int \boldsymbol{d}_j'(y)\mathbf{A}\mathbf{X}'\Delta_n(y)\ dG(y)\}^2 \\ &\le \rho_n \int \|\mathbf{A}\mathbf{X}'\Delta_n(y)\|^2\ dG(y) \end{aligned} \tag{14}$$

and

$$\|\mathbf{Z}_n - \int \mathbf{A}\mathbf{X}'\boldsymbol{\alpha}_n(y)\ f(y)dG(y)\|^2 \le \rho_n \int \|\mathbf{A}\mathbf{X}'\boldsymbol{\alpha}_n\|^2 dG. \tag{15}$$

Moreover,

$$E\int \|\mathbf{A}\mathbf{X}'\boldsymbol{\alpha}_n\|^2\ dG = \int \Sigma_i \|\mathbf{x}_i'\mathbf{A}\|^2\ F_i(1 - F_i)\ dG. \tag{16}$$

Consequently, (v) follows from (6), (7) and (14) whereas (vi) follows from (5.5.5), (6), (15) and (16). Finally, with $\mathscr{D}^{1/2} = \mathbf{A}\mathbf{X}'\mathscr{C}^{1/2}$, $\forall\ \boldsymbol{\theta}$,

$$|k_n(\boldsymbol{\theta}) - \int f\, g dG| = |\boldsymbol{\theta}' \int \mathcal{D}\, g dG\, \boldsymbol{\theta}| = \int \|\boldsymbol{\theta}'\mathcal{D}^{1/2}\|^2\, g dG.$$

Therefore,

$$\sup_{\|\boldsymbol{\theta}\|=1} |k_n(\boldsymbol{\theta}) - \int f\, g dG| \le \int \{\Sigma_i \|\mathbf{A}\mathbf{x}_i\|^2 |f_i - f|\}\, g dG$$
$$\le \rho_n \{\int g^2\, dG\}^{1/2} = o(1), \qquad \text{by (6).} \quad \square$$

**Corollary 5.6b.2.** *Assume that* (1.1.1) *holds with the actual and the modeled d.f.'s of* $\{e_{ni},\ 1 \le i \le n\}$ *equal to* $\{F_{ni},\ 1 \le i \le n\}$ *and* F, *respectively. In addition, assume that* (5.5.1), (5.5.3) – (5.5.7), (5.5.9), (5.5.10) *with* $\mathbf{D} = \mathbf{X}\mathbf{A}$, $H_{ni} \equiv F$, $G_n \equiv G$; (5.5.68) *and* (6) *hold.*
*Then,* (5.5.8) *and* (2) *are satisfied and*

$$\text{(17)} \qquad \mathbf{A}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\left(\int f^2 dG\right)^{-1} \{\hat{\mathbf{Z}}_n + \hat{\mathbf{b}}_n\} + o_p(1)$$

*where*

$$\hat{\mathbf{Z}}_n := \int \mathbf{A}\mathbf{X}'\boldsymbol{\alpha}_n(y)\, d\psi(y) = \mathbf{A}\, \Sigma_i\, \mathbf{x}_{ni}\, [\psi(e_{ni}) - \int \psi(x) dF_{ni}(x)],$$

$$\hat{\mathbf{b}}_n := \int \mathbf{A}\mathbf{X}'\Delta_n(y)\, d\psi(y) = \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni} - F]\, d\psi,$$

*with* $\psi$ *as in* (5.6a.2). □

Consider $\hat{Z}_n$. Note that with $\sigma_{ni}^2 := \operatorname{Var}\{\psi(e_{ni}) | F_{ni}\}$, $1 \le i \le n$,

$$E\, \hat{\mathbf{Z}}_n \hat{\mathbf{Z}}_n' = \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\mathbf{x}_{ni}'\mathbf{A} \cdot \sigma_{ni}^2.$$

One can rewrite

$$\sigma_{ni}^2 = \int\int [F_{ni}(x \wedge y) - F_{ni}(x) F_{ni}(y)]\, d\psi(x) d\psi(y), \qquad 1 \le i \le n.$$

By (5.5.68a), $\psi$ is nondecreasing and bounded. Hence $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$ readily implies that $\max_i \sigma_{ni}^2 \longrightarrow \sigma^2$, $\sigma^2 := \operatorname{Var}\{\psi(e) | F\}$. Moreover, we have the inequality

$$|E\hat{\mathbf{Z}}_n \hat{\mathbf{Z}}_n' - \sigma^2 \mathbf{I}_{p \times p}| \le \sum_i \|\mathbf{A}\mathbf{x}_{ni}\|^2\, |\sigma_{ni}^2 - \sigma^2|.$$

It thus readily follows from the L-F CLT that (5.5.1) implies that $\hat{Z}_n \xrightarrow[d]{} N(0, \sigma^2 \mathbf{I}_{p \times p})$, if $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$. Consequently, we have

**Theorem 5.6b.1.** (*Qualitative Robustness*). *Assume the same setup and conditions as in Corollary* 5.6b.2. *In addition, suppose that*

(18) $$\max_i \|F_{ni} - F\|_\infty = o(1),$$

(19) $$\|A\|_\infty = o(1).$$

*Then, the distribution of* $\hat{\beta}$ *under* $\prod_{i=1}^{n} F_{ni}$ *converges weakly to the degenerate distribution, degenerate at* $\beta$.

**Proof.** It suffices to show that the asymptotic bias is bounded. To that effect we have the inequality

$$\|(\int f^2 dG)^{-1}\, \hat{b}_n\|^2 \le \int \|AX'\Delta\|^2\, dG < \infty, \qquad \text{by (7).}$$

From this, (17), and the above discussion about $\{\hat{Z}_n\}$, we obtain that $\forall\ \eta > 0\ \exists\ K_\eta$ such that $P^n(E_\eta) \longrightarrow 1$, where $P^n$ denotes the probability under $\prod_{i=1}^{n} F_{ni}$ and $E_\eta = \{\|A^{-1}(\hat{\beta} - \beta)\| \le K_\eta\}$. Theorem now follows from this and the elementary inequality $\|\hat{\beta} - \beta\| \le \|A\|_\infty \|A^{-1}(\hat{\beta} - \beta)\|$. □

**Remark 5.6b.2.** The conditions (6) and (18) together need not imply (5.5.7), (5.5.9) and (5.5.10). The condition (5.5.10) is heavily dependent on the rate of convergence in (18). Note that

(20) $$\|\hat{b}_n\|^2 \le \min\{\psi(\infty)\int \|AX'\Delta\|^2 d\psi,\ (\int f^2 dG)\int \|AX'\Delta\|^2 dG\}.$$

This inequality shows that because of (5.5.68), it is possible to have $\|\hat{b}_n\|^2 = O(1)$ even if (7) (or (5.5.10) with $D = XA$) may not be satisfied. However, our general theory requires (7) any way.

Now, with $\varphi = \psi$ or G,

(21) $$\int \|AX'\Delta\|^2 d\varphi = \int \Sigma_i \Sigma_j\, x_i' A A x_j\, \Delta_i\, \Delta_j\, d\varphi \le \int (\Sigma_i \|Ax_i\|\, |\Delta_i|)^2 d\varphi.$$

Thus, if

(22) $$\Sigma_i \|Ax_i\|\ |F_i(y) - F(y)| \le k\, \Delta_n^*(y), \qquad y \in \mathbb{R},$$

where $k$ is a constant and $\Delta_n^*$ is a function such that

$$(23)\qquad \limsup_n \int (\Delta_n^*)^2 d\varphi < \infty,$$

then (7) would be satisfied and in view of (20), $\|\hat{b}_n\| = O(1)$.

Inequality (22) clearly shows that not every sequence $\{F_{ni}\}$ satisfying (6), (18) and (5.5.3) – (5.5.9) with $D = XA$ will satisfy (7). The rate at which $F_{ni} \Rightarrow F$ is crucial for the validity of (7) or (22). □

We now discuss two interesting examples.

**Example 5.6b.1.** $F_{ni} = (1 - \delta_{ni})F + \delta_{ni} F_0$, $1 \le i \le n$. This is the Case 5.5.2. From the Claim 5.5.3, (5.5.5) – (5.5.9) are satisfied by this model as long as (5.5.68) – (5.5.70) and (5.5.1) hold. To see if (6) and (7) are satisfied, note that here

$$\rho_n = \int (\Sigma_i \|Ax_i\|^2 \delta_i |f - f_0\|)^2 dG \le 2 \max_i \delta_i^2 p^2 \cdot [\int (f^2 + f_0^2) dG],$$

and

$$\Sigma_i \|Ax_i\| \, |F_i - F| = \Sigma_i \|Ax_i\| \delta_i \, |F - F_0|.$$

Consequently, here (6) is implied by (5.5.68) for $(f, G)$, $(f_0, G)$ and by (5.5.71), while (7) follows from (5.5.72), (21)–(23) upon taking $\Delta_n^* \equiv |F - F_0|$, provided we additionally assume that

$$(24)\qquad \Sigma_i \|Ax_i\| \delta_i = O(1).$$

There are two obvious choices of $\{\delta_i\}$ that satisfy (24). They are:

$$(25)\qquad (a)\ \ \delta_{ni} = n^{-1/2} \quad \text{or} \quad (b)\ \ \delta_{ni} = p^{-1/2} \|Ax_{ni}\|, \qquad 1 \le i \le n.$$

The gross error models with $\{\delta_i\}$ given by (25b) are more natural than those given by (25a) to linear regression models with unbounded designs. We suggest that in these models, a proportion of contamination one can allow for the $i^{th}$ observation is $p^{-1/2}\|Ax_i\|$. If $\delta_i$ is larger than this in the sense that $\Sigma_i \|Ax_i\| \delta_i \longrightarrow \infty$ then the bias of $\hat{\beta}$ blows up.

Note that if $G$ is a finite measure, $f$ uniformly continuous and $\{\delta_i\}$ are given by (25b) then all the conditions of the above theorem are satisfied by the above $\{F_i\}$ and $F$. Thus we have

**Corollary 5.6b.3.** *Every $\hat{\beta}$ corresponding to a finite measure $G$ is qualitatively robust for $\beta$ against hetroscadastic gross errors at all those $F$'s which have uniformly continuous densities provided $\{\delta_i\}$ are given by* (25b) *and provided* (5.5.1) *and* (19) *hold.* □

**Example 5.6b.2.** Here we consider $\{F_{ni}\}$ given in the Case 5.5.3. We leave it to the reader to verify that one choice of $\{\sigma_{ni}\}$ that implies (7) is to take

$$\sigma_{ni} = \|\mathbf{A}\mathbf{x}_{ni}\|, \qquad 1 \leq i \leq n. \tag{26}$$

One can also verify that in this case, (5.5.68) – (5.5.70), (5.5.75) and (5.5.76) entail the satisfaction of all the conditions of Theorem 5.6b.1. Again, the following corollary holds.

**Corollary 5.6b.4.** *Every* $\hat{\beta}$ *corresponding to a finite measure* G *is qualitatively robust for* $\beta$ *against hetroscedastic scale errors at all those* F's *which have uniformly continuous densities provided* $\{\sigma_{ni}\}$ *are given by* (26) *and provided* (5.5.1) *and* (19) *hold.* □

As an example of a $\sigma$–finite G with $G(\mathbb{R}) = \infty$ that yields a robust estimator, consider $G(y) \equiv (2/3)y$. Assume that the following hold.

(i) F, $F_0$ have continuous densities f, $f_0$; $0 < \int f^2 d\lambda, \int f_0^2\, d\lambda < \infty$.

(ii) $\int F(1-F)\, d\lambda < \infty$. (iii) $\int |F - F_0| d\lambda < \infty$.

Then the corresponding $\hat{\beta}$ is qualitatively robust at F against the heteroscedastic gross errors of Example 5.6b.1 with $\{\delta_{ni}\}$ given by (25b).

Recall, from Remark 5.6a.1, that this $\hat{\beta}$ is also asymptotically efficient at logistic errors. Thus we have a m.d. estimator $\hat{\beta}$ that is asymptotically efficient *and* qualitatively robust at logistic error d.f. against the above gross errors models!!

We leave it to an interested reader to obtain analogues of the above results for $\beta^+$ and $\beta^*$. The reader will find Theorems 5.5.4 and 5.5.6 useful here. □

## 5.6c Locally Asymptotically Minimax Property

In this subsection we shall show that the class of m.d. estimators $\{\beta^+\}$ are locally asymptotically minimax (l.a.m.) in the Hájek – Le Cam sense (Hájek (1972), Le Cam (1972)). In order to achieve this goal we need to recall an inequality from Beran (1982) that gives a lower bound on the local asymptotic minimax risk for estimators of Hellinger differentiable functionals on the class of product probability measures. Accordingly, let $Q_{ni}$, $P_{ni}$ be probability measures on $(\mathbb{R}, \mathscr{B})$, $\mu_{ni}$, $\nu_{ni}$ be a $\sigma$–finite measures on $(\mathbb{R}, \mathscr{B})$ with $\nu_{ni}$ dominating $Q_{ni}$, $P_{ni}$; $q_{ni} := dQ_{ni}/d\nu_{ni}$, $p_{ni} := dP_{ni}/d\nu_{ni}$; $1 \leq i \leq n$. Let $Q^n = Q_{n1} \times \dots \times Q_{nn}$ and $P^n = P_{n1} \times \dots \times P_{nn}$ and $\Pi^n$ denote the class of all n–fold product probability measures $\{Q^n\}$ on $(\mathbb{R}^n, \mathscr{B}^n)$.

Define, for a $c > 0$ and for sequences $0 < \eta_{n1} \longrightarrow 0$, $0 < \eta_{n2} \longrightarrow 0$,

$$\mathscr{H}_n(P^n,c) = \{Q^n \in \Pi^n;\ \Sigma_i \int (q_{ni}^{1/2} - p_{ni}^{1/2})^2 d\nu_{ni} \leq c^2\},$$

$$\mathscr{K}_n(P^n,c,\eta_n) = \{Q^n \in \Pi^n;\ Q^n \in \mathscr{H}_n(P^n,c),\ \max_i \int (q_{ni} - p_{ni})^2 d\mu_{ni} \leq \eta_{n1},$$
$$\max_i \int (q_{ni}^{1/2} - p_{ni}^{1/2})^2 d\nu_{ni} \leq \eta_{n2}\},$$

where $\eta_n' := (\eta_{n1}, \eta_{n2})$.

*DEFINITION* 5.6c.1. A sequence of vector valued functionals $\{S_n: \Pi^n \to \mathbb{R}^p,\ n \geq 1\}$ is Hellinger–(H–) differentiable at $\{P^n \in \Pi^n\}$ if there exists a triangular array of $p \times 1$ random vectors $\{\xi_{ni},\ 1 \leq i \leq n\}$ and a sequence of $p \times p$ matrices $\{A_n,\ n \geq 1\}$ having the following properties:

(i) $\int \xi_{ni} dP_{ni} = 0$, $\int \|\xi_{ni}\|^2 dP_{ni} < \infty$, $1 \leq i \leq n$; $\Sigma_i \int \xi_{ni}\xi_{ni}'\, dP_{ni} \equiv I_{p\times p}$.

(ii) For every $0 < c < \infty$, every sequence $\eta_n \longrightarrow 0$,

$$\sup \| A_n\{S_n(Q^n) - S_n(P^n)\} - 2\, \Sigma_i \int \xi_{ni}\, p_{ni}^{1/2}(q_{ni}^{1/2} - p_{ni}^{1/2})\, d\nu_{ni} \| = o(1)$$

where the supremum is over all $Q^n \in \mathscr{K}_n(P^n,c,\eta_n)$.

(iii) For every $\epsilon > 0$ and every $\alpha \in \mathbb{R}^p$, with $\|\alpha\| = 1$,

$$\Sigma_i \int (\alpha'\xi_{ni})^2 I(|\alpha'\xi_{ni}| > \epsilon)\, dP_{ni} = o(1).$$

Now, let $X_{n1}, \ldots, X_{nn}$ be independent r.v.'s with $Q_{n1}, \ldots, Q_{nn}$ denoting their respective distributions and $\hat{S}_n = \hat{S}_n(X_{n1}, \ldots, X_{nn})$ be an estimator of $S_n(Q^n)$. Let $\mathscr{U}$ be a nondecreasing bounded function on $[0, \infty]$ to $[0, \infty)$ and define the risk of estimating $S_n$ by $\hat{S}_n$ to be

(1) $$R_n(\hat{S}_n, Q^n) = E^n\{\mathscr{U}(\|A_n\{\hat{S}_n - S_n(Q^n)\}\|\},$$

where $E^n$ is the expectation under $Q^n$.

**Theorem 5.6c.1.** *Suppose that* $\{S_n: \Pi^n \to \mathbb{R}^p,\ n \geq 1\}$ *is a sequence of H–differentiable functionals and that the sequence* $\{P^n \in \Pi^n\}$ *is such that*

(2) $$\max_i \int p_{ni}^2\, d\mu_{ni} = O(1).$$

*Then,*

(3) $$\lim_{c\to 0} \liminf_n \inf_{\hat{S}_n} \sup_{Q^n \in \mathscr{K}_n(P^n,c,\eta_n)} R_n(\hat{S}_n, Q^n) \geq E\, \mathscr{U}(\|Z\|)$$

*where* $Z$ *is a* $N(0, I_{pxp})$ *r. v.*

**Sketch of a proof.** This is a reformulation of a result of Beran (1982), pp 425–426. He actually proved (3) with $\mathscr{K}_n(P^n,c,\gamma_n)$ replaced by $\mathscr{H}_n(P^n,c)$ and without requiring (2). The assumption (2) is an assumption on the fixed sequence $\{P^n\}$ of probability measures. Beran's proof proceeds as follows:

Under (i) and (iii), there exists a sequence of probability measures $\{Q^n(\mathbf{h})\}$ such that for every $0 < b < \infty$,

(4) $$\sup_{\|\mathbf{h}\| \leq b} \Sigma_i \int \{q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2} - (1/2)\, \mathbf{h}'\xi_{ni}\, p_{ni}^{1/2}\}^2 d\nu_{ni} = o(1).$$

Consequently,

(5) $$\lim_n \sup_{\|\mathbf{h}\| \leq b} \Sigma_i \int \{q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2}\}^2 d\nu_{ni} = 4^{-1} b^2,$$

and for $n$ sufficiently large, the family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b,\ \mathbf{h} \in \mathbb{R}^p\}$ is a subset of $\mathscr{H}_n(P^n,(b/2))$. Hence, $\forall\ c > 0$, $\forall$ sequence of statistics $\{\hat{S}_n\}$,

(6) $$\liminf_n \inf_{\hat{S}_n} \sup_{Q^n \in \mathscr{H}_n(P^n,c)} R_n(\hat{S}_n, Q^n)$$
$$\geq \liminf_n \inf_{\hat{S}_n} \sup_{\|\mathbf{h}\| \leq 2c} R_n(\hat{S}_n, Q^n(\mathbf{h})).$$

Then the proof proceeds as in Hájek – Le Cam setup for the parametric family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$, under the l.a.n. property of the family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$ with $b = 2c$, which is implied by (4).

Thus (3) would be proved if we verify (6) with $\mathscr{H}_n(P^n,c)$ replaced by $\mathscr{K}_n(P^n,c,\eta_n)$, under the additional assumption (2). That is, we have to show that there exist sequences $0 < \eta_{n1} \to 0$, $0 < \eta_{n2} \to 0$ such that the above family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$ is a subset of $\mathscr{K}_n(P^n,(b/2),\eta_n)$ for sufficiently large $n$. To that effect we recall the family $\{Q^n(\mathbf{h})\}$ from Beran. With $\xi_{ni}$ as in (i) – (iii), let $\xi_{nij}$ denote the jth component of $\xi_{ni}$, $1 \leq j \leq p$, $1 \leq i \leq n$. By (iii) there exist a sequence $\epsilon_n > 0$, $\epsilon_n \downarrow 0$ such that

$$\max_{1 \leq j \leq p} \sum_i \int \xi_{nij}^2\, I(|\xi_{nij}| > \epsilon_n)\, dP_{ni} = o(1).$$

Now, define

$$\xi^*_{nij} := \xi_{nij}\, I(|\xi_{nij}| \le \epsilon_n), \qquad \bar{\xi}_{nij} := \xi^*_{nij} - \int \xi^*_{nij}\, dP_{ni}, \qquad 1 \le j \le p,$$

$$\bar{\xi}_{ni} := (\bar{\xi}_{ni1}, \dots, \bar{\xi}_{nip})', \qquad 1 \le i \le n.$$

Note that

$$(7) \qquad \|\bar{\xi}_{ni}\| \le 2p\epsilon_n, \qquad \int \bar{\xi}_{ni} dP_{ni} = 0, \qquad 1 \le i \le n.$$

For a $0 < b < \infty$, $\|\mathbf{h}\| \le b$, $1 \le i \le n$, define

$$\begin{aligned} q_{ni}(\mathbf{h}) &= (1 + \mathbf{h}'\bar{\xi}_{ni})p_{ni}, & \epsilon_n &< (2bp)^{-1}, \\ &= p_{ni}, & \epsilon_n &\ge (2bp)^{-1}. \end{aligned}$$

Because of (7), $\{q_{ni}(\mathbf{h}),\ \|\mathbf{h}\| \le b,\ 1 \le i \le n\}$ are probability density functions. Let $\{Q_{ni}(\mathbf{h});\ \|\mathbf{h}\| \le b,\ 1 \le i \le n\}$ denote the corresponding probability measures and $Q^n(\mathbf{h}) = Q_{n1}(\mathbf{h})x....xQ_{nn}(\mathbf{h})$.

Now, note that for $\|\mathbf{h}\| \le b$, $1 \le i \le n$,

$$\begin{aligned} \int (q_{ni}(\mathbf{h}) - p_{ni})^2\, d\mu_{ni} &= 0, & \epsilon_n &\ge (2bp)^{-1}, \\ &= \int (\mathbf{h}'\bar{\xi}_{ni})^2 p_{ni}^2\, d\mu_{ni}, & \epsilon_n &< (2bp)^{-1}. \end{aligned}$$

Consequently, since $\epsilon_n \downarrow 0$, $\epsilon_n < (2bp)^{-1}$ eventually, and

$$\sup_{\|\mathbf{h}\| \le b} \max_i \int (q_{ni}(\mathbf{h}) - p_{ni})^2\, d\mu_{ni} \le (2p\epsilon_n)^2\, b^2 \max_i \int p_{ni}^2\, d\mu_{ni} =: \eta_{n1}.$$

Similarly, for a sufficiently large n,

$$\sup_{\|\mathbf{h}\| \le b} \max_i \int (q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2})^2\, d\nu_{ni} \le 2bp\epsilon_n =: \eta_{n2}, \quad \text{say.}$$

Because of (2) and because $\epsilon_n \downarrow 0$, $\max\{\eta_{n1}, \eta_{n2}\} \longrightarrow 0$.

Consequently, for every $b > 0$ and for $n$ sufficiently large, $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \le b\}$ is a subset of $\mathscr{K}_n(P^n,(b/2),\eta_n)$ with the above $\eta_{n1}$, $\eta_{n2}$ and an analogue of (6) with $\mathscr{H}_n(P^n,c)$ replaced by $\mathscr{K}_n(P^n,(b/2),\eta_n)$ holds. The rest is the same as in Beran. □

We shall now show that $\beta^+$ achieves the lower bound in (3). Fix a $\beta \in \mathbb{R}^p$ and consider the model (1.1.1). As before, let $F_{ni}$ be the actual d.f. of $e_{ni}$, $1 \le i \le n$, and suppose we model the errors to be i.i.d F, F symmetric around zero. The d.f. F need not be known. Then the actual and the modeled d.f. of $Y_{ni}$ of (1.1.1) is $F_{ni}(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta})$, $F(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta})$, respectively.

In Theorem 5.6c.1 take $X_{ni} \equiv Y_{ni}$ and $\{Q_{ni}, P_{ni}, \nu_{ni}\}$ as follows:

(8) $$Q_{ni}^{\boldsymbol{\beta}}(Y_{ni} \le \cdot) = F_{ni}(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta}), \;\; P_{ni}^{\boldsymbol{\beta}}(Y_{ni} \le \cdot) = F(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta}),$$
$$\mu_{ni}^{\boldsymbol{\beta}}(\cdot) = G(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta}), \qquad \nu_{ni} \equiv \lambda, \qquad 1 \le i \le n.$$

Also, let $Q_{\boldsymbol{\beta}}^{n} = Q_{n1}^{\boldsymbol{\beta}} x \dots x Q_{nn}^{\boldsymbol{\beta}}$ ; $P_{\boldsymbol{\beta}}^{n} = P_{n1}^{\boldsymbol{\beta}} x \dots x P_{nn}^{\boldsymbol{\beta}}$. The absence of $\boldsymbol{\beta}$ from the sub– or the super– script of a probability measure indicates that the measure is being evaluated at $\boldsymbol{\beta} = \mathbf{0}$. Thus, for example we write $Q^n$ for $Q_{\mathbf{0}}^{n}$ $(= \prod_{i=1}^{n} F_{ni})$ and $P^n$ for $P_{\mathbf{0}}^{n}$, etc. Also for an integrable function g write $\int g$ for $\int g \, d\lambda$.

Let $f_{ni}$, f denote the respective densities of $F_{ni}$, F, w.r.t. $\lambda$. Then $q_{ni}^{\boldsymbol{\beta}}(\cdot) = f_{ni}(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta})$, $p_{ni}^{\boldsymbol{\beta}}(\cdot) = f(\cdot - \mathbf{x}_{ni}'\boldsymbol{\beta})$ and, because of the translation invariance of the Lebesgue measure,

(9) $$\mathscr{H}_n(P_{\boldsymbol{\beta}}^{n}, c) = \{Q_{\boldsymbol{\beta}}^{n} \in \Pi^n \, ; \, \Sigma_i \int \{(q_{ni}^{\boldsymbol{\beta}})^{1/2} - (p_{ni}^{\boldsymbol{\beta}})^{1/2}\}^2 \le c^2\}$$
$$= \{Q^n \in \Pi^n \, ; \, \Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le c^2\} = \mathscr{H}_n(P^n, c).$$

That is the set $\mathscr{H}_n(P_{\boldsymbol{\beta}}^{n}, c)$ does not depend on $\boldsymbol{\beta}$. Similarly,

$$\mathscr{K}_n(P_{\boldsymbol{\beta}}^{n}, c, \eta_n) = \{Q^n \in \Pi^n ; \, Q^n \in \mathscr{H}_n(P^n, c), \, \max_i \int (f_{ni} - f)^2 \, dG \le \eta_{n1},$$
$$\max_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le \eta_{n2}\} = \mathscr{K}_n(P^n, c, \eta_n).$$

Next we need to define the relevant functionals. For $\mathbf{t} \in \mathbb{R}^p$, $y \in \mathbb{R}$, $1 \le i \le n$, define

(10) $$m_{ni}^{+}(y, \mathbf{t}) = F_{ni}(y + \mathbf{x}_{ni}'(\mathbf{t} - \boldsymbol{\beta})) - 1 + F_{ni}(-y + \mathbf{x}_{ni}'(\mathbf{t} - \boldsymbol{\beta})),$$
$$\mathbf{b}_n(y, \mathbf{t}) := \Sigma_i \, \mathbf{A}\mathbf{x}_{ni} \, m_{ni}^{+}(y, \mathbf{t}),$$
$$\mu_n(\mathbf{t}, Q_{\boldsymbol{\beta}}^{n}) \equiv \mu_n(\mathbf{t}, \mathbf{F}) := \int \|\mathbf{b}_n(y, \mathbf{t})\|^2 \, dG(y),$$
$$\mathbf{F}' := (F_{n1}, \cdots, F_{nn}).$$

Now, recall the definition of $\psi$ from (5.6a.2) and let $\mathbf{T}_n(\boldsymbol{\beta}, Q_{\boldsymbol{\beta}}^{n}) \equiv \mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F})$ be defined by the relation

(11) $$\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) := \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X}\int f^2 dG)^{-1} \int \Sigma_i \mathbf{x}_{ni}[F_{ni}(y) - 1 + F_{ni}(-y)]\, d\psi(y).$$

Note that, with $\mathbf{b}_n(y) \equiv \mathbf{b}_n(y, \boldsymbol{\beta})$,

(12) $$\mathbf{A}^{-1}(\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) - \boldsymbol{\beta}) = (\int f^2 dG)^{-1} \int \mathbf{b}_n(y)\, d\psi(y).$$

Some times we shall write $\mathbf{T}_n(\mathbf{F})$ for $\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F})$.

Observe that if $\{F_{ni}\}$ are symmetric around 0, then $\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) = \boldsymbol{\beta} = \mathbf{T}_n(\boldsymbol{\beta}, P^n_{\boldsymbol{\beta}})$. In general, the quantity $\mathbf{A}^{-1}(\mathbf{T}_n(\mathbf{F}) - \boldsymbol{\beta})$ measures the asymptotic bias in $\boldsymbol{\beta}^+$ due to the asymmetry of the errors.

We shall prove the l.a.m. property of $\boldsymbol{\beta}^+$ by showing that $\mathbf{T}_n$ is H-differentiable and that $\boldsymbol{\beta}^+$ is an estimator of $\mathbf{T}_n$ that achieves the lower bound in (3). To that effect we first state a lemma. Its proof follows from Theorem 5.5.4 in the same fashion as that of Lemma 5.6b.1 and Corollary 5.6b.2 from Theorem 5.5.3. Observe that the conditions (5.5.35) and (5.5.11⁺) with $\mathbf{D} = \mathbf{XA}$, respectively, become

(13) $$\int \|\mathbf{b}_n(y)\|^2\, dG(y) = O(1),$$

(14) $$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} \boldsymbol{\theta}'\mathbf{A}\mathbf{X}' \int \boldsymbol{\Lambda}^+ g dG\, \mathbf{X}\mathbf{A}\,\boldsymbol{\theta} \geq \alpha, \qquad \text{for an } \alpha > 0,$$

where $\boldsymbol{\Lambda}^+$ is defined at (5.5.38) and $g$ is as in (5.6a.3).

**Lemma 5.6c.1.** *Assume that* (1.1.1) *holds with the actual d.f.'s of* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$ *and suppose that we model the errors to be i.i.d* $F$, $F$ *symmetric around zero. In addition, assume that* (5.3.8); (5.5.1), (5.5.3), (5.5.4), (5.5.6), (5.5.7), (5.5.9) *with* $\mathbf{D} = \mathbf{XA}$, $G_n \equiv G$; (5.5.68), (5.6a.13), (5.6b.6) *and* (13) *hold. Then* (5.5.8) *and its variant where the argument* $y$ *in the integrand is replaced by* $-y$, (5.5.33), (14) *and the following hold.*

(15) $$\mathbf{A}^{-1}(\boldsymbol{\beta}^+ - \mathbf{T}_n(\mathbf{F})) = -\{2\int f^2\, dG\}^{-1}\, \mathbf{Z}_n^+ + o_p(1), \quad \textit{under } \{Q^n\}.$$

*where*

(16) $$\mathbf{Z}_n^+ = \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, \{\psi(-e_{ni}) - \psi(e_{ni}) - \int m_{ni}^+(y)\, dG(y)\},$$

*with* $m_{ni}^+(y) \equiv m_{ni}^+(y, \boldsymbol{\beta})$ *and* $\psi$ *as in* (5.6a.2). □

Now, define, for an $0 < a < \infty$,

$$\mathscr{M}_n(P^n, a) = \{Q^n \in \Pi^n;\ Q^n = \prod_{i=1}^{n} F_{ni},\ \max_i \int |f_{ni} - f|^r\, dG \longrightarrow 0,\ r = 1, 2,$$
$$\max_i \|F_{ni} - F\|_\infty \longrightarrow 0,\ \int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG \le a^2\}.$$

**Lemma 5.6c.2.** *Assume that* (1.1.1) *holds with the actual d.f.'s of* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$ *and suppose that we model the errors to be i.i.d* F, F *symmetric around zero. In addition, assume that* (5.3.8), (5.5.1), (5.5.68) *and the following hold.*

(17) G *is a finite measure.*

*Then, for every* $0 < a < \infty$ *and sufficiently large* n,

$$\mathscr{M}_n(P^n, a) \supset \mathscr{K}_n(P^n, b_a, \eta_n), \qquad b_a := (4p\alpha)^{-1/2} a, \quad \alpha := G(\mathbb{R}).$$

*Moreover, all assumptions of Lemma* 5.6c.1 *are satisfied.*

**Proof.** Fix an $0 < a < \infty$. It suffices to show that

$$\Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le b_a^2, \qquad n \ge 1, \tag{19}$$

and

$$\text{(a)}\ \max_i \int (f_{ni} - f)^2\, dG \le \eta_{n1}, \qquad n \ge 1, \tag{20}$$

$$\text{(b)}\ \max_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le \eta_{n2}, \qquad n \ge 1,$$

imply all the conditions describing $\mathscr{M}_n(P^n, a)$.

***Claim:*** (19) implies $\int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG \le a^2, \qquad n \ge 1.$

By the C–S inequality,

$$\begin{aligned} |F_{ni}(x) - F(x)|^2 &= \Big|\int_{-\infty}^{x} (f_{ni} - f)\Big|^2 \\ &\le \int_{-\infty}^{x} (f_{ni}^{1/2} - f^{1/2})^2 \cdot \int_{-\infty}^{x} (f_{ni}^{1/2} + f^{1/2})^2 \\ &\le 4 \int (f_{ni}^{1/2} - f^{1/2})^2, \qquad 1 \le i \le n,\ x \in \mathbb{R}. \end{aligned} \tag{21}$$

Hence,

$$\begin{aligned} \int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG &\le \Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|^2 \cdot \Sigma_i \int (F_{ni} - F)^2\, dG \\ &\le 4pa^2 \cdot \Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2, \end{aligned}$$

which proves the Claim.

The finiteness of $G$ together with (21) and (20b) with $\eta_{n2} \longrightarrow 0$ imply that $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$ in a routine fashion. The rest uses (5.5.66), (5.5.67) and details are straightforward. □

Now let $\varphi(y) = \psi(-y) - \psi(y)$, $y \in \mathbb{R}$. Note that $d\psi(-y) \equiv -d\psi(y)$, $d\varphi \equiv -2\, d\psi$, $d\psi = f dG$ and because $F$ is symmetric around 0, $\int \varphi f = 0$. Let

$$\sigma^2 = \mathrm{Var}\{\psi(e)\,|\,F\}, \quad \tau = \int f^2 dG, \quad \rho = (\varphi/\sigma),$$

$$\xi_{ni} \equiv \xi_{ni}(Y_{ni}, \beta) \equiv \mathbf{A}\mathbf{x}_{ni}\, \rho(e_{ni}).$$

Use the above facts to obtain

$$
\begin{aligned}
2\, \Sigma_i \int \xi_{ni}(y, \beta)\, (p^{\beta}_{ni}(y))^{1/2} \{(q^{\beta}_{ni}(y))^{1/2} - (p^{\beta}_{ni}(y))^{1/2}\}^2\, dy \\
&= 2\, \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \int \rho\, f^{1/2}\, (f_{ni}^{1/2} - f^{1/2}) \\
&= \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \{\int \rho\, f_{ni} - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2\} \\
&= -\sigma^{-1}\, \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\{\int [F_{ni} - F]\, d\varphi - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2\}\,. \\
&= \sigma^{-1} \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, \{2 \int [F_{ni} - F]\, f dG - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2\}. \qquad (22)
\end{aligned}
$$

The last but one equality follows from integrating the first term by parts.

Now consider the r.h.s. of (12). Note that because $F$ and $G$ are symmetric around 0,

$$
\begin{aligned}
\int \mathbf{b}_n\, f dG &= \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni}(y) - 1 + F_{ni}(-y)]\, d\psi(y) \\
&= \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni}(y) - F(y) + F_{ni}(-y) - F(-y)]\, d\psi(y) \\
&= 2 \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni} - F]\, f dG. \qquad (23)
\end{aligned}
$$

Recall that by definition $T_n(\beta, P^n_\beta) \equiv \beta$. Now take $A_n$ of (ii) of the H–differentiable requirement to be $\mathbf{A}^{-1}\tau\sigma^{-1}$ and conclude from (18), (22), (23), that

$$
\begin{aligned}
\|A_n\{T_n(\beta, Q^n_\beta) - T_n(\beta, P^n_\beta)\} - \\
- 2\Sigma_i \int \xi_{ni}(y, \beta)(p^{\beta}_{ni}(y))^{1/2}\{(q^{\beta}_{ni}(y))^{1/2} - (p^{\beta}_{ni}(y))^{1/2}\}^2\, dy\, \|
\end{aligned}
$$

$$\leq \|\Sigma_i \mathbf{A}\mathbf{x}_{ni} \int \int \rho \,(f_{ni}^{1/2} - f^{1/2})^2\| \leq \max_i \|\mathbf{A}\mathbf{x}_{ni}\| \cdot \|\rho\|_\infty \cdot b_a^2 = o(1),$$

uniformly for $\{Q^n\} \in \mathscr{K}_n(P^n, b_a, \gamma_n)$.

This proves that the requirement (ii) of the Definition 5.6c.1 is satisfied by the functional $\mathbf{T}_n$ with the $\{\xi_{ni}\}$ given as above. The fact that these $\{\xi_{ni}\}$ satisfy (i) and (iii) of the Definition 5.6c.1 follows from (5.3.8), (5.5.1), (17), (18) and the symmetry of F. This then verifies the H-differentiability of the above m.d. functional $\mathbf{T}_n$.

We shall now derive the asymptotic distribution of $\boldsymbol{\beta}^+$ under any sequence $\{Q^n\} \in \mathscr{M}_n(P^n, a)$, under the conditions of Lemma 5.6c.2. For that reason consider $\mathbf{Z}_n^+$ of (16). Note that under $Q^n$, $(1/2)\mathbf{Z}_n^+$ is the sum of independent centered triangular random arrays and the boundedness of $\psi$ and (5.5.1), imply, via the L–F CLT, that $\mathbf{C}_n^{-1/2}\, \mathbf{Z}_n^+ \xrightarrow[d]{} N(0, \mathbf{I}_{p\times p})$, where

$$\mathbf{C}_n = 4^{-1} E\, \mathbf{Z}_n^+ \mathbf{Z}_n^{+\prime} = \Sigma_i \mathbf{A}\mathbf{x}_{ni}\mathbf{x}_{ni}^{\prime}\mathbf{A}\, \sigma_{ni}^2, \quad \sigma_{ni}^2 = \operatorname{Var}\{\psi(e_{ni})\,|\,F_{ni}\},\ 1 \leq i \leq n.$$

But the boundedness of $\psi$ implies that $\max_i |\sigma_{ni}^2 - \sigma^2| \to 0$, for every $Q^n \in \mathscr{M}_n(P^n, a)$, where $\sigma^2 = \operatorname{Var}\{\psi(e_1)\,|\,F\}$. Therefore $\sigma^{-1}\mathbf{Z}_n^+ \xrightarrow[d]{} N(0, \mathbf{I}_{p\times p})$.

Consequently, from (15),

$$\lim_{c \to 0} \lim_n \sup_{\boldsymbol{\beta}} \sup_{Q^n \in \mathscr{K}_n(P_{\boldsymbol{\beta}}^n, c, \eta_n)} E\{\, \mathscr{U}(\|A_n(\boldsymbol{\beta}^+ - \mathbf{T}_n(\boldsymbol{\beta}, Q_{\boldsymbol{\beta}}^n))\|\,|\,Q_{\boldsymbol{\beta}}^n\} = E\, \mathscr{U}(\|\mathbf{Z}\|).$$

for every bounded nondecreasing function $\mathscr{U}$, where $\mathbf{Z}$ is a $N(0, \mathbf{I}_{p\times p})$ r. v..

This and Lemma 5.6c.2 shows that the sequence of the m.d. estimators $\{\boldsymbol{\beta}^+\}$ achieves the lower bound of (3) and hence is l.a.m. □

**Remark 5.6c.1.** It is an *interesting problem* to see if one can remove the requirement of the finiteness of the integrating measure G in the above l.a.m. result. The l.a.m. property of $\{\hat{\boldsymbol{\beta}}\}$ can be obtained in a similar fashion. For an alternative definition of l.a.m. see Millar (1984) where, among other things, he proves the l.a.m. property, in his sense, of $\{\hat{\boldsymbol{\beta}}\}$ for $p = 1$.

***A problem:*** To this date an appropriate extension of Beran (1978) to the model (1.1.1) does not seem to be available. Such an extension would provide asymptotically fully efficient estimators at every symmetric density with finite Fisher information and would also be l.a.m. □

Note: The contents of this chapter are based on the works of Williamson (1979, 1982), Koul (1979, 1980, 1984, 1985a,b), Koul and DeWet (1983), Basaw and Koul (1988) and Dhar (1991a, b). □□