

Robust Estimation of the Conditional Median Function at Elliptical Models

Christophe Croux^a, Catherine Dehon^a,
Peter J. Rousseeuw^{b,*}, and Stefan Van Aelst^{b,†}

^a ECARES, Université Libre de Bruxelles, CP-139, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium.

^b Dept. of Mathematics and Computer Science, Universitaire Instelling Antwerpen, Universiteitsplein 1, B-2610 Antwerpen, Belgium.

Abstract: In this note we study the problem of estimating the parameters of the conditional median function at elliptical models. For this we use positive-breakdown estimators of multivariate location and scatter, and obtain influence functions and asymptotic variances of the resulting slope and intercept. We also consider a technique to gain efficiency by artificially increasing the dimension.

Key words: Elliptical Model, Median, Regression, Robustness.

*Corresponding author. Tel: +3238202412, fax: +3238202421.

E-mail address: Peter.Rousseeuw@uia.ua.ac.be.

†Stefan Van Aelst is Research Assistant with the FWO Belgium.

1 Introduction

Let x be a p -variate explanatory variable and y a univariate response variable. Denote $z = (x^t, y)$ and suppose that z is distributed according to an elliptically symmetric distribution H . The density h of H can then be expressed as

$$h(x) = \frac{1}{(\det \Sigma)^{1/2}} g((z - \mu)^t \Sigma^{-1} (z - \mu)) \quad (1)$$

where g is a real-valued function, $\mu \in \mathbb{R}^{p+1}$ is called the location parameter, and the parameter $\Sigma \in \text{PDS}(p+1)$ is called the scatter matrix. Here $\text{PDS}(p+1)$ is the set of positive definite symmetric matrices of size $p+1$. We can decompose μ and Σ as

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix},$$

and similarly for estimators $\hat{\mu}_n$ and $\hat{\Sigma}_n$ of μ and Σ . It follows from Proposition 13.1 of Bilodeau and Brenner (1999, page 208) that the conditional distribution of y given x is symmetric with median

$$m(x) := \text{med}(y|x) = \alpha + \beta^t x \quad (2)$$

where

$$\beta = \Sigma_{xx}^{-1} \Sigma_{xy} \quad \text{and} \quad \alpha = \mu_y - \beta^t \mu_x. \quad (3)$$

Since we did not assume that H has a first moment, the conditional mean $E[y|x]$ need not exist. Note that (2) can be rewritten as a linear regression equation

$$y = \alpha + \beta^t x + \varepsilon \quad (4)$$

where the distribution of the error term $\varepsilon := y - m(x)$ is symmetric about zero. In general this distribution depends on x , so (4) is a heteroscedastic regression model. The only H for which the distribution of ε does not depend on x is the multivariate normal (see e.g. Fang, Kotz and Ng 1990, page 106). In that case ε has a univariate gaussian distribution with constant scale, hence (4) becomes homoscedastic.

For any elliptical distribution H it follows from (2) that the conditional median function $m(x)$ is linear, so it is natural to estimate it by

$$\hat{m}_n(x) := \hat{\alpha}_n + \hat{\beta}_n^t x \quad (5)$$

with

$$\hat{\beta}_n = \hat{\Sigma}_{n,xx}^{-1} \hat{\Sigma}_{n,xy} \quad \text{and} \quad \hat{\alpha}_n = \hat{\mu}_{n,y} - \hat{\beta}_n^t \hat{\mu}_{n,x}. \quad (6)$$

In this way, every estimator $(\hat{\mu}_n, \hat{\Sigma}_n)$ of multivariate location and scatter leads to an immediate estimate of the conditional median function. When $E_H(|Y|) < \infty$ it holds that $m(x) = E[y|x]$, so then $\hat{m}_n(x)$ also estimates the conditional mean function.

The classical estimator of (μ, Σ) is the empirical mean and covariance matrix. For an i.i.d. sample z_1, \dots, z_n from H , with $z_i = (x_i^t, y_i)$, they are given by

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n z_i \quad \text{and} \quad \hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{\mu}_n)(z_i - \hat{\mu}_n)^t.$$

The corresponding $\hat{\alpha}_n$ and $\hat{\beta}_n$ as given by (6) can then easily be shown to coincide with the least squares regression estimator of y on x . This is a highly nonrobust procedure, since it is known that the least squares estimator can become completely corrupted by even a single outlier. Therefore we prefer to insert a robust estimator of multivariate location and scatter instead. Many such estimators have been proposed (see Maronna and Yohai 1998 for an overview). We will require them to be affine equivariant and asymptotically normal. Maronna and Morgenthaler (1986) inserted M-estimators of multivariate location and scatter, and showed that the resulting estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ have all the desired equivariance properties. However, it is known that the breakdown value of multivariate M-estimators goes down like $1/p$.

In this paper we will focus on estimators of (μ, Σ) that have a high breakdown value, such as the Minimum Covariance Determinant estimator (MCD) of Rousseeuw (1984). For a finite sample of observations z_1, \dots, z_n in \mathbb{R}^{p+1} , the MCD is determined by selecting that subset $\{z_{i_1}, \dots, z_{i_h}\}$ of size h , with $1 \leq h \leq n$, which minimizes the generalized variance (which is the determinant of the covariance matrix computed from the subset) among all possible subsets of size h . The location estimator is then defined as

$$\hat{\mu}_n = \frac{1}{h} \sum_{j=1}^h z_{i_j}$$

and the scatter estimator by

$$\hat{\Sigma}_n = c_{p+1} \frac{1}{h} \sum_{j=1}^h (z_{i_j} - \hat{\mu}_n)(z_{i_j} - \hat{\mu}_n)^t \quad (7)$$

where c_{p+1} is a consistency factor. The choice $h = [(n + p + 2)/2] \approx 0.5n$ yields the highest possible breakdown value. The commonly preferred default value is $h \approx 0.75n$, yielding a better compromise between efficiency and breakdown value.

In Section 2 of this paper we compute the influence function of the estimators of α and β based on robust estimators of location and scatter. In Section 3 we see that the efficiency of the estimators depends on the number of explanatory variables. This fact can be exploited to slightly increase the efficiency by artificially augmenting the dimension

2 Influence Function

The influence function (IF) measures the effect of a small amount of outliers on an estimator (Hampel et al, 1986). Before deriving influence functions we need to define the functionals corresponding to $\hat{\alpha}_n$ and $\hat{\beta}_n$. Let G be an arbitrary $(p + 1)$ dimensional distribution and $(T(G), C(G)) \in \mathbb{R}^{p+1} \times PDS(p + 1)$ the location/scatter functional evaluated at G . For instance, an S-estimator (see Rousseeuw and Leroy 1987, page 263) can be defined as a functional, and for finite samples we put $\hat{\mu}_n := T(H_n)$ and $\hat{\Sigma}_n := C(H_n)$ where H_n is the empirical distribution function of the data. We now decompose

$$T(G) = \begin{bmatrix} T_x(G) \\ T_y(G) \end{bmatrix} \quad \text{and} \quad C(G) = \begin{bmatrix} C_{xx}(G) & C_{xy}(G) \\ C_{yx}(G) & C_{yy}(G) \end{bmatrix}.$$

The functionals of interest are then

$$b(G) = C_{xx}(G)^{-1}C_{xy}(G) \tag{8}$$

and

$$a(G) = T_y(G) - b(G)^t T_x(G). \tag{9}$$

By definition, a Fisher-consistent location/scatter functional satisfies $T(H) = \mu$ and $C(H) = \Sigma$ at the model distribution H , yielding

$$b(H) = \beta \quad \text{and} \quad a(H) = \alpha.$$

At the empirical distribution function H_n we put $\hat{\beta}_n = b(H_n)$ and $\hat{\alpha}_n = a(H_n)$.

The influence function of b at the model distribution H is defined as

$$\text{IF}(z, b, H) = \left. \frac{\partial}{\partial \varepsilon} b((1 - \varepsilon)H + \varepsilon\Delta_z) \right|_{\varepsilon = 0}$$

where Δ_z represents a Dirac measure putting all its mass at $z = (x^t, y) \in \mathbb{R}^{p+1}$, and $\text{IF}(z, a, H)$ is defined analogously. Since we require that T and C be affine equivariant, we may suppose without loss of generality that $\mu = 0$ and $\Sigma = I_{p+1}$ which yields $\alpha = 0$ and $\beta = 0$. If T and C possess an influence function, then by Lemma 2 of Croux and Haesbroeck (2000) we know that two real-valued functions γ and δ exist such that

$$\text{IF}(z, C, H) = \gamma(z^t z)z^t z - \delta(z^t z)I_{p+1}. \quad (10)$$

Furthermore, there exists a third real-valued function ζ such that

$$\text{IF}(z, T, H) = \zeta(z^t z)z. \quad (11)$$

We will express the influence function of the slope and intercept in terms of γ and ζ .

Proposition 1 *At the model distribution H with $\mu = 0$ and $\Sigma = I_{p+1}$ it holds that*

$$\text{IF}(z, b, H) = \gamma(z^t z)xy \quad \text{and} \quad \text{IF}(z, a, H) = \zeta(z^t z)y$$

with $z = (x^t, y)$.

Proof: This follows from applying the chain rule to (8) and (9) and filling in (10) and (11), because

$$\begin{aligned} \text{IF}(z, b, H) &= C_{xx}(H)^{-1}\text{IF}(z, C_{xy}, H) + \text{IF}(z, C_{xx}^{-1}, H)C_{xy}(H) \\ &= \text{IF}(z, C_{xy}, H) = \gamma(z^t z)xy \end{aligned}$$

and

$$\begin{aligned} \text{IF}(z, a, H) &= \text{IF}(z, T_y, H) - \text{IF}(z, b, H)^t T_x(H) - b(H)^t \text{IF}(z, T_x, H) \\ &= \text{IF}(z, T_y, H) = \zeta(z^t z)y \end{aligned}$$

where the standardizations $C(H) = I_{p+1}$ and $T(H) = 0$ were used. \square

For multivariate M-estimators the functions γ , δ and ζ can be obtained from Huber (1981, page 226). For S-estimators of multivariate location and scatter the functions γ , δ and ζ are given by Lopuhaä (1989), and for the MCD estimator they can be retrieved from Butler, Davies, and Jhun (1993) and Croux and Haesbroeck (1999a). Lopuhaä (1999) gives expressions for the reweighted versions of these estimators. It then follows directly from Proposition 1 that the influence functions of the slope and intercept based on these estimators will be

bounded, by filling in their functions γ and ζ . For instance, for the $(p+1)$ -dimensional MCD estimator with trimming proportion τ (i.e. with coverage $h \approx (1 - \tau)n$) we obtain simple explicit formulas for γ and ζ :

$$\gamma_{\text{MCD}}(t) = \frac{I(t \leq \sqrt{q_\tau})}{\Gamma(\frac{p+1}{2} + 2, \frac{q_\tau}{2})} \quad \text{and} \quad \zeta_{\text{MCD}}(t) = \frac{I(t \leq \sqrt{q_\tau})}{\Gamma(\frac{p+1}{2} + 1, \frac{q_\tau}{2})} \quad (12)$$

where $q_\tau = \chi_{p+1, 1-\tau}^2$ and $\Gamma(u, v) = \Gamma(u)^{-1} \int_0^v e^{-t} t^{u-1} dt$ is the incomplete gamma function. The influence functions of a and b based on the MCD approach can thus be seen as truncated versions of the Least Squares influence functions (where $\gamma \equiv \zeta \equiv 1$). In the center of the distribution both $\text{IF}(z, b, \text{MCD})$ and $\text{IF}(z, a, \text{MCD})$ are similar to their classical counterparts, but observations far from the center (i.e. with large $z^t z$) will receive a zero influence.

3 Efficiency Considerations

It is shown in Butler, Davies, and Jhun (1993) that if the function g in (1) has a continuous first derivative which is strictly negative, then the MCD satisfies

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} N(0, \text{ASV}(T, H))$$

with asymptotic (co-)variance $\text{ASV}(T, H) = E_H[\text{IF}(z, T, H)\text{IF}(z, T, H)^t]$, and similarly for the MCD scatter estimator $\hat{\Sigma}$. Since $\hat{\beta}$ is a differentiable function of the elements of $\hat{\Sigma}$, the estimator $\hat{\beta}$ will also be asymptotically normal (see e.g. Barndorff-Nielsen and Cox 1989, page 47) with $\text{ASV}(b, H) = E_H[\text{IF}(z, b, H)\text{IF}(z, b, H)^t]$, and similarly for \hat{a} . This leads to

Proposition 2 *At the model distribution H with $\mu = 0$ and $\Sigma = I_{p+1}$ and density h given by (1) such that g has a continuous first derivative which is strictly negative, we have*

$$\text{ASV}(b, H) = \left(\int \gamma^2(z^t z) y^2 x_1^2 dH(x, y) \right) I_p$$

and

$$\text{ASV}(a, H) = \int \zeta^2(z^t z) y^2 dH(x, y).$$

Formal regularity conditions for the asymptotic normality of S-estimators for location and scatter can be found in Corollary 5.1 of Lopuhaä (1989). With these regularity conditions Proposition 2 also holds for S-estimators.

Table 1: Asymptotic variances of the MCD-based slope and intercept for different values of p .

p	1	2	3	4	5	10	15
b	6.1350	4.2918	3.5176	3.0864	2.8120	2.2141	1.9919
a	2.4788	2.1456	1.9819	1.8828	1.8155	1.6536	1.5859

From (10) and (11) one can easily verify that $ASV(b, H) = ASV(C_{xy}, H)$ and $ASV(a, H) = ASV(T_y, H)$. Therefore, the efficiency of the slope estimator equals that of the off-diagonal elements of the scatter matrix estimator, and the intercept estimator $\hat{\alpha}_n$ inherits the efficiency of the location estimator.

Expressions for the asymptotic variances of the MCD-based slope and intercept in the Gaussian case can be derived from Proposition 2 together with (12). With trimming portion τ we have that $ASV(b, H) = (1/\Gamma(\frac{p+1}{2} + 2, \frac{q\tau}{2}))I_p$ and $ASV(a, H) = 1/\Gamma(\frac{p+1}{2} + 1, \frac{q\tau}{2})$. Table 1 lists the asymptotic variances of slope and intercept for $\tau = 0.25$ and different values of p . (The matrix $ASV(b, H)$ is represented by one of its diagonal elements.) In Figure 1 we see how these asymptotic variances decrease with the number of explanatory variables. Note that for the classical estimator these asymptotic variances are independent of the dimension and equal to one.

Since the asymptotic variance of the median function estimate $m(x) = a + b^t x$ can be expressed as a function of the asymptotic variances of the parameter estimates a and b , the impact of adding extra variables on the efficiency of the estimate $m(x)$ is similar to the impact on the efficiency of the parameter estimates.

The decrease of asymptotic variance seen in Figure 1 has also been observed for multivariate location estimators such as the spatial median (Brown, 1983), the Oja median (Oja and Niinimaa, 1985) and S-estimators of location (Rocke, 1996).

A surprising consequence of this increasing efficiency is that adding explanatory variables will increase the precision of the slope and intercept estimators. These extra variables may even be randomly generated. This technique is illustrated in the following simulation experiment, where we are interested in estimating the conditional median function for (x, y) distributed according to a bivariate normal distribution H :

$$\text{median}(y|x) = \alpha + \beta x. \tag{13}$$

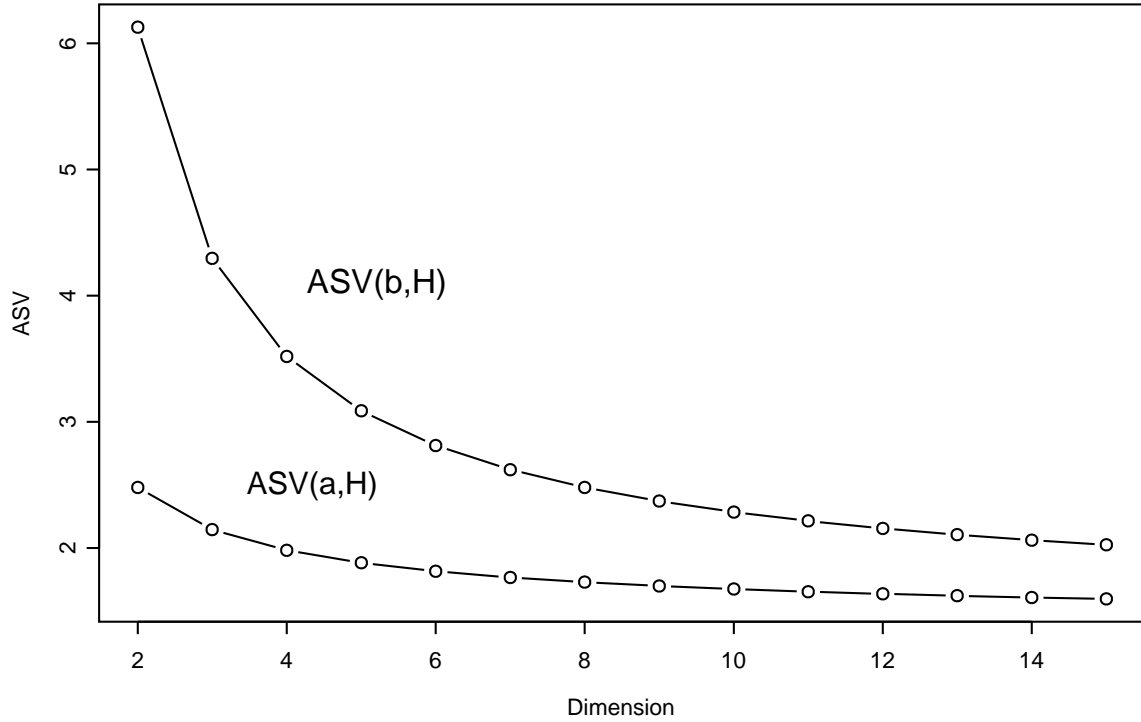


Figure 1: $ASV(b, H)$ and $ASV(a, H)$ for the slope and intercept functionals based on the MCD with trimming proportion 25%, as a function of the dimension $p + 1$ and for $H = N(0, I_{p+1})$.

Table 2: Simulated bias and MSE for the slope and intercept estimators of the conditional median function at the bivariate normal model for several sample sizes. The estimates are based on the MCD estimator and on the MCD+1 and MCD+3 procedures.

n		$\hat{\beta}$			$\hat{\alpha}$		
		MCD	MCD+1	MCD+3	MCD	MCD+1	MCD+3
20	bias	0.0298	0.0136	0.0170	-0.0027	-0.0203	0.0072
	n*MSE	5.9315	6.0863	8.2312	2.7085	3.2405	4.4914
50	bias	0.0107	-0.0078	-0.0120	-0.0039	-0.0107	-0.0059
	n*MSE	5.7326	5.2264	4.4659	2.5204	2.5204	2.5607
100	bias	-0.0057	-0.0010	0.0053	0.0007	0.0010	0.0027
	n*MSE	5.5339	4.5309	3.3948	2.3747	2.2689	2.1411
200	bias	0.0086	0.0079	0.0056	-0.0024	-0.0017	0.0000
	n*MSE	6.7878	4.8495	3.4645	2.3401	2.1985	1.8538
∞	bias	0	0	0	0	0	0
	n*MSE	6.1350	4.2918	3.0864	2.4788	2.1456	1.8828

A total of $m = 1000$ samples of size n was generated from $H = N(0, I_2)$ (we take $\Sigma = I_2$ w.l.o.g. due to equivariance properties). For each sample, estimates $\hat{\alpha}_n^j$ and $\hat{\beta}_n^j$ based on the MCD with $h = \lfloor \frac{3}{4}n \rfloor$ were computed, using the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999). The simulated bias is defined as $\frac{1}{m} \sum_{j=1}^m \hat{\beta}_m^j - \beta$, and the mean squared error (MSE) as $\frac{1}{m} \sum_{j=1}^m (\hat{\beta}_m^j - \beta)^2$. Similar definitions apply for the intercept. From Table 2 we see that the biases are negligible, while the efficiencies at finite samples are relatively close to their asymptotic counterparts for $n \geq 50$.

The same simulation has been repeated for what will be called the MCD+1 procedure. Here we generated an extra variable $v \sim N(0, 1)$ independently of x and y . By estimating the parameters of the equation

$$\text{median}(y|x) = \text{median}(y|x, v) = \alpha_1 + \beta_1 x + \beta_2 v$$

using the three-dimensional MCD location/scatter estimator we obtain $\hat{\beta}_1$ and $\hat{\beta}_2$. We set $\hat{\alpha} := \hat{\alpha}_1$ (since the parameter α in (13) equals α_1) and $\hat{\beta} := \hat{\beta}_1$, and we discard $\hat{\beta}_2$. The estimator MCD+3 is defined similarly, and requires computation of the MCD in 5 dimensions. From Table 2 we see that adding the variable v did decrease the mean squared error at not-too-small finite samples ($n \geq 50$ for the slope and $n \geq 100$ for the intercept). This decrease is more pronounced for the slope $\hat{\beta}$ than for the intercept $\hat{\alpha}$. The efficiency gain is not only observed by simulation for finite n , but (for the last row in Table 2) also follows from the exact asymptotic variances obtained from Table 1. We see in Figure 1 that the gain in efficiency for the intercept is indeed less substantial than for the slope. Note that the breakdown value of MCD, MCD+1 and MCD+3 remains unchanged at 25%.

The question which arises is what price we have paid for this increase in efficiency. One would expect an increase in the bias of the estimators. Therefore, in each generated sample we replaced 20% of the observations by outliers (x_i, y_i) generated from the contaminating distribution $N_2((5, 5)^t, \text{diag}(0.1, 0.1))$. The resulting biases and MSEs are given in Table 3. We again see a decrease of the MSE, but now only for $n \geq 100$ for the slope and $n \geq 200$ for the intercept. It is remarkable that all the biases remain negligible, and that the MSEs under contamination turn out to be even better than the MSEs when no outliers are present. This is because the MCD estimator has a redescending influence function so that far away outliers are harmless, and therefore applying the MCD with $1 - \tau = 75\%$ to data with 20% of far outliers is like applying the MCD with $1 - \tau = 75\%/80\% = 93.75\%$ to a (smaller) data set without outliers.

Table 3: As in Table 2, but with 20% of extreme outliers.

		$\hat{\beta}$			$\hat{\alpha}$		
n		MCD	MCD+1	MCD+3	MCD	MCD+1	MCD+3
20	bias	0.0264	0.0362	0.1279	-0.0003	-0.0087	-0.0046
	n*MSE	2.6388	3.4920	6.2751	1.8250	2.1900	4.2271
50	bias	0.0027	-0.0052	-0.0023	0.0008	0.0007	0.0031
	n*MSE	2.4485	2.2801	2.6368	1.7628	1.7381	1.7960
100	bias	-0.0024	-0.0032	-0.0028	0.0000	0.0011	0.0016
	n*MSE	2.5243	2.1791	2.1530	1.6162	1.5386	1.6116
200	bias	0.0014	-0.0003	0.0004	-0.0033	-0.0024	-0.0002
	n*MSE	2.3635	2.0785	1.8405	1.5195	1.4361	1.4022

Table 4: As in Table 2, but with 20% of intermediate outliers.

		$\hat{\beta}$			$\hat{\alpha}$		
n		MCD	MCD+1	MCD+3	MCD	MCD+1	MCD+3
20	bias	0.5997	0.4823	0.3837	0.1327	0.1435	0.1542
	n*MSE	10.6688	9.1956	9.0498	4.4554	4.6984	6.1127
50	bias	0.6248	0.5229	0.4142	0.1329	0.1505	0.1677
	n*MSE	22.8656	17.5709	12.8190	5.9116	5.2684	5.1440
100	bias	0.5987	0.5095	0.4155	0.1446	0.1686	0.1845
	n*MSE	38.7552	29.2909	20.4300	6.6277	6.3930	6.2680
200	bias	0.5957	0.5153	0.4241	0.1517	0.1697	0.1834
	n*MSE	74.1416	56.5352	39.1476	9.5074	9.4878	9.4194

It was shown by Croux and Haesbroeck (1999b) that intermediate outliers are more harmful than extreme outliers, so we conducted a third experiment where the outliers are generated from the distribution $N_2((\sqrt{q/2}, \sqrt{q/2})^t, \text{diag}(0.1, 0.1))$, with $q = \chi_{2,0.75}^2$ the upper 25% quantile of the χ_2^2 distribution. In Table 4 we see that the results for the slope parameter improve when the dimension is increased, but for the intercept the MSEs remain similar.

To conclude this simulation experiment, we could say that adding artificially generated random variables increases the efficiency of the estimator for sample sizes which are not too small, without loss of robustness. This increase in efficiency is substantial for small p , but for problems with moderate to high p the effect will practically disappear, since already at the asymptotic level the gain becomes very small (see Figure 1). It is surprising to see that the efficiency increase (which is in line with earlier results for estimation of multivariate location and covariance) costs very little in terms of contamination bias. Of course, one should not forget that the computational effort increases with every added variable

4 References

- Barndorff-Nielsen, O.E., and Cox, D.R., 1989. *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.
- Bilodeau, M., and Brenner, D., 1999. *Theory of Multivariate Statistics*, Springer Verlag, New York.
- Brown, B.M., 1983. Statistical uses of the spatial median, *J. R. Statist. Soc. B* 45, 25-30.
- Butler, R.W., Davies, P.L. and Jhun, M., 1993. Asymptotics for the minimum covariance determinant estimator, *Ann. Statist.* 21, 1385–1400.
- Croux, C., and Haesbroeck, G., 1999a. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *J. of Multivariate Anal.* 71, 161–190.
- Croux, C., and Haesbroeck, G., 1999b. Maxbias curves of robust scale estimators based on subranges, Preprint, University of Brussels (ULB), submitted.
- Croux, C. and Haesbroeck, G., 2000. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, *Biometrika*, to appear in September.
- Fang, K., Kotz, S., and Ng, K., 1990. *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., 1986. Robust Statistics: The Approach based on Influence Functions, Wiley, New York.
- Huber, P.J., 1981. Robust Statistics, Wiley, New York.
- Lopuhaä, H.P., 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.* 17, 1662–1683.
- Lopuhaä, H.P., 1999. Asymptotics of reweighted estimators of multivariate location and scatter, *Ann. Statist.* 27, 1638-1665.
- Maronna, R.A., and Morgenthaler S., 1986. Robust regression through robust covariances, *Commun. Statist.-Theor. Meth.* 15, 1347–1365.
- Maronna, R.A., and Yohai V.J., 1998. Robust estimation of multivariate location and scatter, in: S. Kotz, C. Read, and D. Banks, eds., *Encyclopedia of Statistical Sciences Update Volume 2*, Wiley, New York, pp. 589–596.
- Oja, H., and Niinimaa, A., 1985. Asymptotic properties of the generalized median in the case of multivariate normality, *J. R. Statist. Soc. B* 47, 372-377.
- Rocke, D.M., 1996. Robustness properties of S-estimators of multivariate location and shape in high dimension, *Ann. Statist.* 24, 1327-1345.
- Rousseeuw, P.J., 1984. Least median of squares regression, *J. Amer. Statist. Assoc.* 79, 871–880.
- Rousseeuw, P.J., and Leroy, A.M., 1987. Robust Regression and Outlier Detection, Wiley, New York.
- Rousseeuw, P.J., and Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator, *Technometrics* 41, 212–223.