Ann. Inst. Statist. Math.
Vol. 48, No. 1, 97–110 (1996)

FUZZY WEIGHTED SCALED COEFFICIENTS IN SEMI-PARAMETRIC MODEL*

JONG-WUU WU[1], JIAHN-BANG JANG[2] AND TZONG-RU TSAI[2]

[1] *Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan 25137, R.O.C.*
[2] *Graduate School of Statistics, National Chengchi University, 64, 2nd Section, Chi-nan Rd., Taipei, Taiwan 11623, R.O.C.*

(Received May 23, 1994; revised April 26, 1995)

Abstract. In general, the regressor variables are stochastic, Duan and Li (1987, *J. Econometrics*, **35**, 25–35), Li and Duan (1989, *Ann. Statist.*, **17**, 1009–1052) have been shown that under very general design conditions, the least squares method can still be useful in estimating the scaled regression coefficients of the semi-parametric model $Y_i = Q_1(\alpha + \beta X_i; \epsilon_i)$, $i = 1, 2, \dots, n$. Here α is a constant, β is a $1 \times p$ row vector, X_i is a $p \times 1$ column vector of explanatory variables, ϵ_i is an unobserved random error and Q_1 is an arbitrary unknown function. When the data set (X_i, Y_i), $i = 1, 2, \dots, n$, contains one or several outliers, the least squares method can not provide a consistent estimator of the scaled coefficients β. Therefore, we suggest the "fuzzy" weighted least squares method to estimate the scaled coefficients β for the data set with one or several outliers. It will be shown that the proposed "fuzzy" weighted least squares estimators are $\sqrt{n}$-consistent and asymptotically normal under very general design condition. Consistent measurement of the precision for the estimator is also given. Moreover, a limited Monte Carlo simulation and an example are used to study the practical performance of the procedures.

Key words and phrases: Least squares estimator, semi-parametric model, outlier, asymptotic normality, fuzzy weighted least squares estimator, Monte Carlo simulation.

1. Introduction

In many practical situations, it is important to study the relationship of the major response variable Y to the other factors $X^T = (X_1, X_2, \dots, X_p)$. To examine such relationship one usually uses a regression model in which Y has a distribution that depends on X. In many applications, parametric model is the best approximation to the true model, however, the search for an appropriate model is not easy. This is particularly true when the sample size is limited and

* This research partially supported by the National Science Council, R.O.C.

the justification of the use of the parametric model is difficult. Under such situations, nonparametric smoothing techniques are frequently regarded as promising alternatives. The nonparametric methods break down quickly when the number of explanatory variables becomes large. To overcome the disadvantage, many statisticians and economists suggest the estimation procedures for the scaled coefficients of a semi-parametric model. The semi-parametric model is

$$Y_i = Q_1(\alpha + \beta X_i; \epsilon_i), \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where Q_1 is an arbitrary unknown function, α is a finite constant, β is a $1 \times p$ row vector of coefficient, X_i is a $p \times 1$ column vector of explanatory variables, and ϵ_i is an unobservable random error term, $i = 1, 2, \ldots, n$.

For the model (1.1), Q_1 is unknown, Brillinger (1982) showed that the ordinary least squares method provide useful estimates for the scaled regression coefficients if the regressors are stochastic and jointly Gaussian or deterministic and quasi-Gaussian. Duan and Li (1987), Li and Duan (1989) also investigated conditions on the behavior of regressors such that consistent estimators can be developed for the scaled coefficients of the regression function of the dependent variable.

In some practical conditions, the data set from the population has one or several outliers, (see Rousseeuw and Levoy (1987), Rousseeuw (1984)), we can use the optimal fuzzy clustering procedure proposed by Van Cutsem and Gath (1993) to detect outliers. Therefore, the ordinary least squares method can not provide consistent estimators for the scaled regression coefficients of the model (1.1). In this paper, for the data set with outliers, we suggest an estimation procedure for the scaled coeffficients of a semi-parametric model. Thus, the semi-parametric model is replaced by the model

$$Y_{ji} = Q(\alpha + \beta X_{ji}; \epsilon_{ji}), \qquad i = 1, 2, \ldots, n_j, \quad j = 1, 2, \ldots, q, \tag{1.2}$$

where Q is an arbitrary unknown function, α is a finite constant, β is a $1 \times p$ row vector of coefficients, X_{ji} is a $p \times 1$ column vector of explanatory variables, ϵ_{ji} is an unobservable random error term, q is the group number, $\sum_{j=1}^{q} n_j = n$, $q > 1$ and $n_j \ll n_1$, $j = 2, 3, \ldots, q$. Define n_1 to be the size of main group.

For example, if the data set has only one outlier, then $q = 2$, $n_2 = 1$ and $n_1 = n - 1$ in the model (1.2). Even when the data set has a few outliers, we can use the optimal fuzzy clustering procedure proposed by Van Cutsem and Gath (1993) to classify outliers to some groups.

In this paper, we assume that the regressors are stochastic. We use the weighted least squares method to construct a consistent estimator for the scaled coefficients β, and the optimal fuzzy clustering procedure to provide the weights of the weighted least squares method.

Throughout this paper, our goal is to estimate the slope vector β when the link function Q is unknown. In this circumstance, the intercept α can not be identified, and the slope vector can only be identified up to a multiplicative scalar. Therefore, we can at most estimate β up to an unknown constant of proportionality. Such estimate is useful in testing $H_0 : \beta A = 0$, A is a known $p \times k$ matrix, or estimating the ratios of slope coefficients. These ratios measure the relative importance of the regressor variates of interest. To estimate the main features of Y, such as $E(Y \mid X)$ for the data set with one or several outliers, our estimation procedure can provide a useful tool for reducing the dimension of the explanatory variable X to one dimension.

Our consistent estimator is asymptotically normally distributed under some conditions. The asymptotic variance-covariance matrices can also be shown to be consistently estimated, and hence the precision of the estimate can be evaluated and test statistics can be derived. Section 2 discusses the estimation procedure and introduce the optimal fuzzy clustering procedure. Basically, suppose that for the stochastic regressors in each group have jointly elliptically symmetric distribution (Fang *et al.* (1990)) (or satisfy Condition (C1) of Subsection 2.1), then the weighted least squares method can still help to estimate the scaled coefficient β. Asymptotic results of the proposed estimator and its variance estimator are derived in Section 3. Section 4 provides simulation results of the estimates under two regression models, and one practical example. In the small sample size, the simulation performances of the estimator are still very robust.

2. Estimation and the optimal fuzzy procedure

2.1 *Estimation procedure*

Suppose our set-up (Y_i, X_i), $i = 1, 2, \ldots, n$, are independent but not all identically distributed random vectors with one or several outliers. Therefore by the optimal fuzzy clustering procedure in Subsection 2.2, we can divide the data set into some groups. Suppose there are q groups. Given the random observations (Y_{ji}, X_{ji}), $i = 1, 2, \ldots, n_j$, and $n_j \ll n_1$, $j = 2, 3, \ldots, q$, and $\sum_{j=1}^{q} n_j = n$, with the assumption that for each j the unobserved ϵ_{ji}, $i = 1, 2, \ldots, n_j$ are i.i.d. random vectors, we assume $E(Y_{ji} \mid X_{ji}) = g(\alpha + \beta X_{ji})$, where g is a real-valued function (that is, the means of the observed Y_{ji} depend on X_{ji} only through βX_{ji}). When $q = 1$, if the regressors X^T are Gaussian variables with mean μ^T and variance-coverance matrix Σ, and g is an almost differentiable function satisfying $E|g'(\alpha + \beta X)| < \infty$, then applying Stein's (1981) identity we have, for some constant θ_0,

$$E[g(\alpha + \beta X)(X - \mu)^T \Sigma^{-1}] = \theta_0 \beta \tag{2.1}$$

where

$$\theta_0 = \mathrm{Cov}\{g(\alpha + \beta X), \alpha + \beta X\} / \mathrm{var}(\alpha + \beta X).$$

This result was essentially first obtained by Brillinger (1982). Moveover, Brillinger (1982) also applied this result to show that the ordinary least squares method may provide consistent estimators of β up to a constant of proportionality. Result (2.1) was later provided to be still valid by Duan and Li (1987) and Li and Duan (1989) even we only assume

CONDITION (C1). $E(WX \mid \beta X)$ is linear in βX for all linear combinations WX. Particular multivariate distributions satisfying (C1) include all elliptically symmetric distributions (Fang *et al.* (1990)). In this paper, according to the set-up of data with one or several outliers, assuming $q > 1$ and applying Theorem 1 in Duan and Li (1987), we can prove the whole basis of our "fuzzy" weighted least squares estimator which is the conclusion of the following lemma.

LEMMA 2.1. *For each j, let X_j^T have a multivariate distribution with mean vector μ_j^T and variance-covariance matrix Σ_j. Assume Condition* (C1) *is satisfied. Suppose Y_j is a random variable depending on X_j such that $E(Y_j \mid X_j) = g(\alpha + \beta X_j)$, where the real-valued function g satisfies $E[g(\alpha + \beta X_j){X_j}^T] < \infty$. There exist some finite constants θ_j such that*

$$E[Y_j(X_j - \mu_j)^T {\Sigma_j}^{-1}] = \theta_j \beta, \qquad j = 1, 2, \ldots, q.$$

If the variance-covariance matrix Σ_j, $j = 1, 2, \ldots, q$, are different, and under

CONDITION (C2). $n_j \ll n_1$, $w_{ji} = o(1)$, $i = 1, 2, \ldots, n_j$, $j = 2, 3, \ldots, q$, and for some constants d_j satisfy $n \sum_{i=1}^{n_j} {w_{ji}}^2 = d_j + o(1)$, $j = 1, 2, \ldots, q$, $\sum_{j=1}^{q} \sum_{i=1}^{n_j} w_{ji} = 1$, we can easily show

$$\hat{\beta}_* = \sum_{j=1}^{q} \sum_{i=1}^{n_j} w_{ji} Y_{ji} (X_{ji} - \bar{X}_w)^T \left\{ \sum_{j=1}^{q} \sum_{i=1}^{n_j} w_{ji} (X_{ji} - \bar{X}_w)(X_{ji} - \bar{X}_w)^T \right\}^{-1} \tag{2.2}$$

to be a consistent estimator of $\beta_{**} = \sum_{i=1}^{n_1} w_{1i}\theta_1\beta = \theta_{**}\beta (\theta_{**} = \sum_{i=1}^{n_1} w_{1i}\theta_1)$, where $\bar{X}_w = \sum_{j=1}^{q} \sum_{i=1}^{n_j} w_{ji} X_{ji}$, w_{ji} as weights from the optimal fuzzy procedure, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, q$. Additionally, if all the variance-covariance matrices Σ_j, $j = 1, 2, \ldots, q$ are the same, but the means μ_j, $j = 1, 2, \ldots, q$ are different, then under Condition (C2) we can also show $\hat{\beta}_*$ to be a consistent estimator of $\beta_* = \sum_{j=1}^{q} \sum_{i=1}^{n_j} w_{ji}\theta_j\beta = \theta_*\beta$. The asymptotic normality of $\sqrt{n}(\hat{\beta}_* - \beta_*)$ and $\sqrt{n}(\hat{\beta}_* - \beta_{**})$ is established in Section 3. In addition, we shall propose consistent estimators for the asymptotic variance-covariance matrix. In Subsection 2.2, we will discuss the weights w_{ji}, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, q$ from the optimal fuzzy procedure proposed by Bezdek (1973), Gath and Geva (1989).

2.2 *The optimal fuzzy clustering procedure*

This algorithm is derived from the combination of the fuzzy k-means algorithm proposed by Bezdek (1973) and the fuzzy maximum likelihood estimation proposed by Gath and Geva (1989). Fuzzy k-means algorithm is derived from the solution which minimizes the following fuzzy version least squares function:

$$J(X, V) = \sum_{j=1}^{n} \sum_{i=1}^{K} u_{ij}^{\ell} d^2(X_j, V_i), \qquad K \le n,$$

where u_{ij} is the degree of membership of X_j in i-th cluster, X_j is the j-th p-dimensional data vector, V_i is the i-th cluster center, $d^2(X_j, V_i)$ is the Euclidean distance between X_j and V_i, K is the number of clusters. The weighting exponent for u_{ij}, $\ell(\geq 1)$, controls the "fuzziness" of the resulting clusters. A theoretical basis for an optimal choice of ℓ is so far not available by Bezdek (1981), based on empirical grounds ℓ was chosen to be equal to 2. The steps of fuzzy clustering proposed by Gath and Geva (1989) are following:

1. Carry out unsupervised tracking of the initial set of cluster centers, $(V_1, V_2, \ldots, V_K)$.

2. Calculate the weighted matrix U (with entries u_{ij}) according to

$$u_{ij} = \{1/d^2(X_j, V_i)\}/\left\{\sum_{k=1}^{K}(1/d^2(X_j, V_k))\right\}.$$

3. Calculate the new set of cluster centers $(\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_K)$, where

$$\hat{V}_i = \left\{\sum_{j=1}^{n}(u_{ij})^2 X_j\right\} \Big/ \left\{\sum_{j=1}^{n}(u_{ij})^2\right\}$$

and update U to $\hat{u}_{ij}$, according to Step 2.

4. If $\max_{i,j}[|u_{ij} - \hat{u}_{ij}|] < \epsilon$ stop, else goto Step 3, where ϵ between 0 and 1.

For Step 1, we need two statistics to do unsupervised fuzzy partition. In order to decide the initial subgroups and the optimal numbers of clusters in the data set, Gath and Geva (1989) proposed the performance measures—fuzzy hypervolume (F_{HV}) and average partition density (D_{PA}).

Fuzzy hypervolume (F_{HV}) is defined by

$$F_{HV} = \sum_{i=1}^{K}[\det(F_i)]^{1/2}$$

where

$$F_i = \left\{\sum_{j=1}^{n} h(i \mid X_j)(X_j - V_i)(X_j - V_i)^T\right\} \Big/ \left\{\sum_{j=1}^{n} h(i \mid X_j)\right\},$$

$h(i \mid X_j)$ is the posterior probability (the probability of selecting the i-th cluster given the j-th data vector). When $\ell = 2$, $h(i \mid X_j)$ is similar to u_{ij}.

Average partition density (D_{PA}) is defined by

$$D_{PA} = (1/K)\left\{\sum_{i=1}^{K} S_i/\{[\det(F_i)]^{1/2}\}\right\}$$

where $S_i(= \sum_{j:X_i \in \chi_i} u_{ij})$ is called to be sum of central members, and

$$\chi_i = \{X_j \mid (X_j - V_i)^T F_i^{-1}(X_j - V_i) < 1\}.$$

Gath and Geva (1989) proposed that the optimal numbers of clusters in the data set is determined by K which is corresponding to global minimum in F_{HV} or global maximum in D_{PA}.

If we already get an optimal cluster result, we can choice the normalized degree of membership corresponding to the main classified cluster as weights for (2.2).

3. Asymptotic normality and measurement of precision

First, we devote to establish the asymptotic normality of $\sqrt{n}(\hat{\beta}_* - \beta_*)$. Given the set-up of data in Section 2, and the assumption that for each j the regressors X_{ji} has mean μ_{ji} and variance-covariance matrix Σ, the asymptotic distribution of $\sqrt{n}(\hat{\beta}_* - \beta_*)$ is equivalent to that of

$$S_n = \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}\{(Y_{ji} - EY_j) - \beta_*(X_{ji} - \mu_j)\}(X_{ji} - \mu_j)^T\Sigma^{-1}. \tag{3.1}$$

Since S_n is the average of independent random variables, by the Slutsky theorem and the Theorem B of Serfling ((1980), p. 30), we estabilish the assertion of Theorem 3.1. The asymptotic variance-covariance matrix is also stated in Theorem 3.1 with detailed analysis.

THEOREM 3.1. *Assume conditions* (C1), (C2) *and* (C3): $E[Y_jX_j^T] < \infty$, *for each* $j = 1, 2, \ldots, q$, *are satisfied, then*

$$\sqrt{n}(\hat{\beta}_* - \beta_*) \xrightarrow{d} MVN(0, \Sigma_*), \qquad as \quad n \to \infty$$

provided the variance-covariance matrix Σ_* *exists. Here the matrix* Σ_* *is*

$$\Sigma_* = \Sigma^{-1}\sum_{j=1}^{q} d_j E\{[(Y_j - EY_j) - \beta_*(X_j - \mu_j)]^2(X_j - \mu_j)(X_j - \mu_j)^T\}\Sigma^{-1}.$$

PROOF. Set

$$\hat{\Sigma}_w = \left\{\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}(X_{ji} - \bar{X}_w)(X_{ji} - \bar{X}_w)^T\right\}.$$

Then

$$\begin{aligned}
\hat{\beta}_* &= \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}Y_{ji}(X_{ji} - \bar{X}_w)^T\hat{\Sigma}_w^{-1} \\
&= \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}Y_{ji}(X_{ji} - \mu_j)^T\hat{\Sigma}_w^{-1} + \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}Y_{ji}(\mu_j - \bar{X}_w)^T\hat{\Sigma}_w^{-1} \\
&= \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}(Y_{ji} - EY_j)(X_{ji} - \mu_j)^T\Sigma^{-1} \\
&\quad + \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}(Y_{ji} - EY_j)(X_{ji} - \mu_j)^T(\hat{\Sigma}_w^{-1} - \Sigma^{-1})
\end{aligned}$$

$$+\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} EY_j (X_{ji}-\mu_j)^T \hat{\Sigma}_w^{-1} + \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} Y_{ji} (\mu_j - \bar{X}_w)^T \hat{\Sigma}_w^{-1}$$

$$= \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}(Y_{ji}-EY_j)(X_{ji}-\mu_j)^T\Sigma^{-1}$$

$$+\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} E\{(Y_{ji}-EY_j)(X_{ji}-\mu_j)^T\}(\hat{\Sigma}_w^{-1}-\Sigma^{-1}) + o_p(n^{-1/2}).$$

(By the Slutsky theorem, the Theorem B of Serfling ((1980), p. 30), and (C2))

Set $PB = \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} E\{(Y_{ji}-EY_j)(X_{ji}-\mu_j)^T\}(\hat{\Sigma}_w^{-1}-\Sigma^{-1})$

$$PB = \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} E\{(Y_{ji}-EY_j)(X_{ji}-\mu_j)^T\}\hat{\Sigma}_w^{-1}(\Sigma-\hat{\Sigma}_w)\Sigma^{-1}$$

$$= \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji} E\{(Y_{ji}-EY_j)(X_{ji}-\mu_j)^T\}\Sigma^{-1}(\Sigma-\hat{\Sigma}_w)\Sigma^{-1} + o_p(n^{-1/2})$$

(By the Slutsky theorem, the Theorem B of Serfling ((1980), p. 30), (C2), and $\bar{X}_w \xrightarrow{p} \mu_1$)

$$= -\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}\theta_1\beta(\hat{\Sigma}_w-\Sigma)\Sigma^{-1} + o_p(n^{-1/2})$$

$$= -\beta_*(\hat{\Sigma}_w-\Sigma)\Sigma^{-1} + o_p(n^{-1/2})$$

$$= -\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}\beta_*(X_{ji}-\bar{X}_w)(X_{ji}-\bar{X}_w)^T\Sigma^{-1} + \beta_* + o_p(n^{-1/2})$$

$$= -\sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}\beta_*(X_{ji}-\mu_j)(X_{ji}-\mu_j)^T\Sigma^{-1} + \beta_* + o_p(n^{-1/2}),$$

therefore,

$$\hat{\beta}_* = \sum_{j=1}^{q}\sum_{i=1}^{n_j} w_{ji}\{(Y_{ji}-EY_j)-\beta_*(X_{ji}-\mu_j)\}(X_{ji}-\mu_j)^T\Sigma^{-1} + \beta_* + o_p(n^{-1/2}),$$

then

$$\sqrt{n}(\hat{\beta}_*-\beta_*) \xrightarrow{d} MVN(0,\Sigma_*), \qquad \text{as} \quad n\to\infty,$$

where

$$\Sigma_* = \Sigma^{-1}\sum_{j=1}^{q} d_j E\{[(Y_j-EY_j)-\beta_*(X_j-\mu_j)]^2(X_j-\mu_j)(X_j-\mu_j)^T\}\Sigma^{-1}.$$

The proof is completed. □

For the estimator $\hat{\beta}_*$ to be of practical use, one needs an estimator of its asymptotic variance-covariance matrix. Therefore, a moment estimator of Σ_* can be defined by

$$\hat{\Sigma}_* = \hat{\Sigma}_w^{-1} \sum_{j=1}^{q} (d_j/n_j)$$
$$\cdot \sum_{i=1}^{n_j} \{[(Y_{ji} - \bar{Y}_j) - \hat{\beta}_*(X_{ji} - \bar{X}_j)]^2 (X_{ji} - \bar{X}_j)(X_{ji} - \bar{X}_j)^T\} \hat{\Sigma}_w^{-1}.$$

Under the conditions of Theorem 3.1, $\hat{\Sigma}_*$ can be shown to be asymptotically consistent for Σ_*. Based on this measurement of precision, hypothesis testing on $H_0 : \beta A = 0$, A is a $p \times k$ matrix of full rank $k \leq p$, can be done by using the usual Wald statistic. The test statistic is $n(\hat{\beta}_* A)(A\hat{\Sigma}_* A^T)^{-1}(\hat{\beta}_* A)^T$ and its limiting distribution is χ^2 with k degrees of freedom.

Secondly, given the set-up of data in Section 2, and the assumption that for each j, the regressors X_{ji} has mean μ_j and variance-covariance matrix Σ_j (different variance-covariance matrix), the asymptotic representation similar to (3.1) for $\sqrt{n}(\hat{\beta}_* - \beta_{**})$ can be analogously derived. Thus $\sqrt{n}(\hat{\beta}_* - \beta_{**})$ is equivalent to that of

$$T_n = \sum_{i=1}^{n_1} w_{1i}\{(Y_{1i} - EY_1 - \beta_{**}(X_{1i} - \mu_1))\}(X_{1i} - \mu_1)^T \Sigma_1^{-1}, \tag{3.2}$$

where $\beta_{**} = \sum_{i=1}^{n_1} w_{1i}\theta_1\beta$. Then, $\hat{\beta}_*$ is a $\sqrt{n}$-consistent estimator of β_{**} stated formally in Theorem 3.2.

THEOREM 3.2. *Given Conditions* (C1), (C2) *and* $E[Y_1 {X_1}^T] < \infty$, *then*

$$\sqrt{n}(\hat{\beta}_* - \beta_{**}) \xrightarrow{d} MVN(0, \Sigma_{**}), \qquad as \quad n \to \infty$$

provided the variance-covariance matrix Σ_{**} exists. Here the matrix Σ_{**} is

$$\Sigma_{**} = \Sigma_1^{-1} d_1 E\{[(Y_1 - EY_1) - \beta_{**}(X_1 - \mu_1)]^2 (X_1 - \mu_1)(X_1 - \mu_1)^T\} \Sigma_1^{-1}.$$

Moreover, we can show that given the conditions in Theorem 3.2, Σ_{**} can be estimated consistently by

$$\hat{\Sigma}_{**} = \hat{\Sigma}_w^{-1}(d_1/n_1)$$
$$\cdot \sum_{i=1}^{n_1} \{[(Y_{1i} - \bar{Y}_1) - \hat{\beta}_*(X_{1i} - \bar{X}_1)]^2 (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)^T\} \hat{\Sigma}_w^{-1}.$$

Table 1. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (4, 7)$ in *Model* (I).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44743	0.89432	0.50031
	(0.581)	(0.146)	(0.716)
	(0.363)	(0.091)	
20	0.44732	0.89437	0.50015
	(0.169)	(0.042)	(0.302)
	(0.154)	(0.038)	
30	0.44718	0.89444	0.49995
	(0.101)	(0.025)	(0.156)
	(0.080)	(0.020)	

Table 2. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (1, 5)$ in *Model* (I).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44732	0.89437	0.50015
	(0.425)	(0.106)	(0.663)
	(0.340)	(0.085)	
20	0.44720	0.89443	0.49998
	(0.141)	(0.035)	(0.244)
	(0.126)	(0.031)	
30	0.44714	0.89446	0.49990
	(0.090)	(0.022)	(0.151)
	(0.077)	(0.019)	

Note: The unit in parentheses is 10^{-5}.

4. Simulation studies, example and final remarks

In order to study the finite sample properties of the "fuzzy" weighted least squares estimator $\hat{\beta}_*$, a Monte Carlo experiment has been done in which 300 samples of different sizes n were generated from various populations for dependent variable. While it is not possible to completely characterize the sampling behaviour of the estimator, our main purpose is to suggest that the proposed estimator can be feasible and can perform well in finite samples. Here, we only discuss the case which (X_i, Y_i), $i = 1, 2, \ldots, n$ with same variance-covariance matrix and different means. Alternatively, in our unpublished technical report, (X_i, Y_i), $i = 1, 2, \ldots, n$

Table 3. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (11, 3)$ in *Model* (I).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44723	0.89442	0.50003
	(0.123)	(0.031)	(0.289)
	(0.148)	(0.037)	
20	0.44719	0.89444	0.49997
	(0.049)	(0.012)	(0.128)
	(0.066)	(0.016)	
30	0.44723	0.89442	0.50002
	(0.038)	(0.010)	(0.077)
	(0.040)	(0.010)	

Table 4. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (15, 13)$ in *Model* (I).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44725	0.89441	0.50005
	(0.253)	(0.063)	(0.565)
	(0.290)	(0.073)	
20	0.44719	0.89444	0.49997
	(0.110)	(0.027)	(0.240)
	(0.123)	(0.031)	
30	0.44723	0.89442	0.50003
	(0.075)	(0.019)	(0.131)
	(0.067)	(0.017)	

Note: The unit in parentheses is 10^{-5}.

with different means and different variance-covariance matrices, the estimator $\hat{\beta}_*$ is also very robust in the finite samples.

Two additive models were investigated in our experiment. *Model* (I) considered $Y = 50 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$ where $(\beta_1, \beta_2) = (50, 100)$ and the true standardizable parameter $(\beta_{1*}, \beta_{2*}) = (0.44721, 0.89443)$. The regressors (X_{1i}, X_{2i}) are independent of error ϵ_i, $i = 1, 2, \ldots, n$. For $i = 1, 2, \ldots, n-2$, (X_{1i}, X_{2i}) have jointly multivariate normal distribution with mean $(\mu_1, \mu_2) = (6.0, 8.0)$ and variance and covariances: $\sigma_{11} = 7$, $\sigma_{22} = 8$, $\sigma_{12} = 4$ and for $i = n-1$, n, (X_{1i}, X_{2i}) have jointly multivariate normal distribution with mean (μ_1, μ_2) (where $(\mu_1, \mu_2) = (4, 7)$ or $(1, 5)$ or $(11, 3)$ or $(15, 13)$) and variance and covariances: $\sigma_{11} = 7$, $\sigma_{22} = 8$, $\sigma_{12} = 4$. Random error ϵ_i has standard normal distribution. *Model* (II) is basi-

Table 5. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (4, 7)$ in *Model* (II).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44758	0.89424	0.50052
	(1.923)	(0.483)	(2.285)
	(1.159)	(0.290)	
20	0.44740	0.89433	0.50027
	(0.556)	(0.139)	(0.982)
	(0.501)	(0.125)	
30	0.44718	0.89440	0.49995
	(0.101)	(0.025)	(0.156)
	(0.080)	(0.020)	

Table 6. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (1, 5)$ in *Model* (II).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44736	0.89434	0.50022
	(1.397)	(0.345)	(2.197)
	(1.134)	(0.280)	
20	0.44716	0.89445	0.49992
	(0.461)	(0.115)	(0.805)
	(0.413)	(0.103)	
30	0.44708	0.89449	0.49982
	(0.299)	(0.075)	(0.492)
	(0.251)	(0.063)	

Note: The unit in parentheses is 10^{-5}.

cally the same as the first model except that we define ϵ_i has logistic distribution with mean zero and variance $\pi^2/3$. It is noted that the logistic distribution can be obtained as a mixture of extreme value distributions.

The Monte Carlo estimates of the means of the estimators $\hat{\beta}_{*i}$ for β_{*i} and the ratio estimators $\hat{\gamma} = \hat{\beta}_{*1}/\hat{\beta}_{*2}$ for the scaled invariant parameter $\gamma = \beta_1/\beta_2 = 0.5$ are given in Table 1 to Table 8. Furthermore, the sample variances $S^2(\hat{\beta}_{*i})$ of the estimators $\hat{\beta}_{*i}$ and the Monte Carlo means of the variance estimators $\hat{\sigma}^2(\beta_{*i})$ based on the results in Section 3 and the estimated mean squares error were also calculated in order to investigate the behaviour of the measurements of precision.

We find the experiment results encouraging. Basically speaking, when the

Table 7. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (11, 3)$ in *Model* (II).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44725	0.89440	0.50006
	(0.404)	(0.101)	(0.921)
	(0.473)	(0.118)	
20	0.44717	0.89445	0.49995
	(0.160)	(0.040)	(0.415)
	(0.213)	(0.053)	
30	0.44724	0.89442	0.50003
	(0.125)	(0.031)	(0.256)
	(0.131)	(0.033)	

Table 8. Estimated means of $\hat{\beta}_*$ and $\hat{\gamma}$, the sample variance $\hat{S}^2(\beta_*)$ (in 1st parentheses), the Monte Carlo means of the variance $\hat{\sigma}^2(\beta_*)$ (in 2nd parentheses), and means of squares errors (in parentheses) of $\hat{\gamma}$ for $(\mu_1, \mu_2) = (15, 13)$ in *Model* (II).

n	$\hat{\beta}_{*1}$	$\hat{\beta}_{*2}$	$\hat{\gamma}$
10	0.44726	0.89440	0.50007
	(0.827)	(0.206)	(1.813)
	(0.931)	(0.233)	
20	0.44719	0.89444	0.49997
	(0.355)	(0.089)	(0.774)
	(0.398)	(0.099)	
30	0.44726	0.89440	0.50007
	(0.246)	(0.062)	(0.435)
	(0.223)	(0.056)	

Note: The unit in parentheses is 10^{-5}.

sample size is small, the biases of the estimators $\hat{\beta}_*$ and $\hat{\gamma}$ are almost negligible for *Models* (I) and (II). The improvement of the bias reduction by increasing sample size is significant. The sample variance estimators $S^2(\hat{\beta}_{*i})$ and the Monte Carlo means of the variance estimators $\hat{\sigma}^2(\beta_{*i})$ behave satisfactorily for different sample size under both *Models* (I) and (II). Therefore, in the small sample size, the simulation performances of the estimator are very robust.

Additionally, to illustrate the ideas presented, we have applied the proposed estimate $\hat{\beta}_*$ to the raw data on Wood Specific Gravity from Rousseeuw (1984), which is given in Table 9.

An optimal selection at $k = 2$ for the fuzzy optimal clustering procedure

Table 9. Modefied data on Wood Specific Gravity. u_{1j} be the degree of membership of main group from optimal cluster result by using the fuzzy clustering procedure.

index	X_1	X_2	X_3	X_4	X_5	Y	u_{1j}
1	0.5730	0.1059	0.4650	0.5380	0.8410	0.5340	0.85891
2	0.6510	0.1356	0.5270	0.5450	0.8870	0.5350	0.95306
3	0.6060	0.1273	0.4940	0.5210	0.9200	0.5700	0.93508
*4	0.4370	0.1591	0.4460	0.4230	0.9920	0.4500	0.01617
5	0.5470	0.1135	0.5310	0.5190	0.9150	0.5480	0.89736
*6	0.4440	0.1628	0.4290	0.4110	0.9840	0.4310	0.02067
7	0.4890	0.1231	0.5620	0.4550	0.8240	0.4810	0.62758
*8	0.4130	0.1673	0.4180	0.4300	0.9780	0.4230	0.02603
9	0.5360	0.1182	0.5920	0.4640	0.8540	0.4750	0.77239
10	0.6850	0.1564	0.6310	0.5640	0.9140	0.4860	0.84630
11	0.6640	0.1588	0.5060	0.4810	0.8670	0.5540	0.86623
12	0.7030	0.1335	0.5190	0.4840	0.8120	0.5190	0.82961
13	0.6530	0.1395	0.6250	0.5190	0.8920	0.4920	0.87608
14	0.5860	0.1114	0.5050	0.5650	0.8890	0.5170	0.96400
15	0.5340	0.1143	0.5210	0.5700	0.8890	0.5020	0.88729
16	0.5230	0.1320	0.5050	0.6120	0.9190	0.5080	0.77675
17	0.5800	0.1249	0.5460	0.6080	0.9540	0.5200	0.85767
18	0.4480	0.1028	0.5220	0.5340	0.9180	0.5060	0.47615
*19	0.4170	0.1687	0.4050	0.4150	0.9810	0.4010	0.03883
20	0.5280	0.1057	0.4240	0.5660	0.9090	0.5680	0.62754

Note: "*" represent outlier.

shows that the data are contaminated with outlying observations and discerns observations 4, 6, 8 and 19 as outlying observations (Van Cutsem and Gath (1993)). Therefore, we use the "fuzzy" weighted least squares method to estimate the scaled regression coefficient β_*. Here the weights for (2.2) are from the normalized degree of membership corresponding to the main classified cluster; see Table 9. So, the "fuzzy" weighted least squares estimate is

$$\hat{\beta}_* = (0.3408, -0.7762, -0.4122, -0.2163, 0.4250)$$

and the variance estimate is

$$\hat{\sigma}^2(\beta_*) = (0.0048, 0.0988, 0.0030, 0.0060, 0.0134).$$

And the signs of the estimate $\hat{\beta}_{*i}$ are the same as those of estimate of the least median of squares regression proposed by Rousseeuw and Leroy (1987). Therefore, the "fuzzy" weighted least squares estimate can also provide a good initial estimate for the regression parameters in the least median of squares regression.

In the practical example, we can use the general Wald statistics proposed by Section 3, $\chi^2 = n(\hat{\beta}_* A)(A^T \hat{\Sigma}_* A)^{-1}(\hat{\beta}_* A)^T$ to test $H_0 : \beta_1 = \beta_2 = \beta_3 =$

$\beta_4 = \beta_5 = 0$. Here let A is the identity matrix. Therefore, given $\alpha = 0.05$, $\chi^2 = 110.7 > \chi^2_{(5,0.05)} = 11.071$. So, we can reject H_0 and conclude that at least one of the explanatory variables is related to the dependent variable.

Finally, according to our simulation study, we have seen that the proposed method for analyzing the data set with one or several outlier is reliable. Particularly, we have found the estimator $\hat{\beta}_*$ is easy to compute.

Acknowledgements

We thank the associate editor and the referee for helpful comments.

References

Bezdek, J. C. (1973). Fuzzy mathematics in pattern classification, Ph.D. Thesis, Cornell University, Ithaca.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York.

Brillinger, D. R. (1982). A generalized linear model with 'Gaussian' regressor variables, *A Festchrift for Erich L. Lehmann on Honor of His Sixty-fifth Birthday* (eds. P. J. Bickel, K. Doksum and J. L. Hodges), 97–114, Wadsworth International Group, Belmont, California.

Duan, N. and Li, K.-C. (1987). Distribution-free and link-free estimation for the sample selection model, *J. Econometrics*, **35**, 25–35.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York.

Gath, I. and Geva, A. (1989). Unsupervised optimal fuzzy clustering, *IEEE Trans. on Patterns Analysis and Machine Intelligence*, **11**, 773–781.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation, *Ann. Statist.*, **17**, 1009–1052.

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.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Stein, C. M. (1981). Estimation of the mean of multivariate normal distributions, *Ann. Probab.*, **12**, 361–379.

Van Cutsem, B. and Gath, I. (1993). Detection of outliers and robust estimation using fuzzy clustering, *Comput. Statist. Data Anal.*, **15**, 47–61.