ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 162, 2002

*Maciej Górkiewicz**

ESTIMATION OF MEASUREMENT ERROR USING LOCAL SAMPLING AND JOINT NONPARAMETRIC LINEARISATION

ABSTRACT. This paper presents how to use the near neighbours technique in aim to transform a given data set (Z, X, Y^T) of size N into a set of $J \approx N$ local samples (Z, X), with restrictions on minimal number K of members in each local sample and on maximal difference of Y^T inside each local sample, where Z plays role of an outcome, X is an independent variable, and $Y^T = (Y_1, \ldots, Y_L)$ is a vector of L supplementary continuous variables. Then the procedure for non-parametric joint linearisation of an obtained set of local samples was proposed. The whole proposed method was applied to estimation of models with standard deviation of measurements as outcome Z and measured value as independent variable X. The paper was inspired by difficulties with estimation of the measurement error, which often occur in medicine, if accuracy of a measurement procedure depends on some properties of patient. Nevertheless, the proposed approach seems to be more general. It can be useful in many analyses of observational studies, which aim to estimate a family of the functions, preferable the linear ones, instead a single multivariate model.

Key words: near neighbours, local sampling, linearisation, measurement error.

I. INTRODUCTION

Let us consider a given set of N individuals. Suppose that the vector (Z, X, Y^T) of some continuous variables was observed on each individual, where Z is an outcome, X is an independent variable, and $Y^T = (Y_1, \ldots, Y_L)$ is a vector of L supplementary variables. The problem arises if the multivariate regression approach cannot be applied, because an outcome Z cannot be considered as a function of the continuous covariates Y^T, so the regression cannot be described with single function $Z = f(X, Y^T)$. The conventional idea how this problem can be overcome is as follows. On the beginning in the space of continuous

* Dr, Institute of Public Health, Collegium Medicum, Jagiellonian University of Cracow.

covariates set of representative points $Y_1^T, Y_2^T, ..., Y_J^T$ was chosen. Then for each point Y^T = idem a separate local sample (*Z*, *X*) was drawn and local regression $Z = Z_j(X) \mid Y_j^T$, $j = 1, 2, ..., J$ was estimated. Finally, the relations between functions $Z_j(X)$ were investigated and mapped into space of covariates Y^T (Domański and Pruska 2000). In practice, however, we rarely have opportunity to get a sufficient number of data (*Z*, *X*) for each local sample. For this reason in the paper the nearest neighbours method was examined as tool, which enables us to drawn a needed number *J* of local samples of needed size *K*, with neglected differences of the covariates inside each local sample, from given whole sample of size *N* near to *J*. It is obvious, that this procedure cannot enlarge an initial quantity of information in the data, which remains correspond to size $N < J*K$. A conventional number of independent local samples was assumed near to $N/K < J$.

It is known, that estimation and testing of hypothesis about parameters in a non linear regression model has continued to present formidable problems. The difficulty lies mainly in the fact that the statistical methods of inference which have optimal properties in linear models are not optimal in non-linear models (Crowder and Hand 1990; Davidian and Giltinan 1995). For this reason in the paper the non-parametric iterative procedure for joint linearisation of all set of local samples was proposed. The iterative transformations are justified with mini-max criterion of the consistency of the resulting linear estimators.

It is known that the standard deviation of measurements for fixed measured value and fixed covariates can be easy estimated by a few repeated measurements (Bland and Altman 1986). In such way one can obtain needed sample of data: estimated outcome *Z* (standard deviation) – independent variable *X* (measured value) – covariates Y^T. In practice three procedures for estimation of standard deviation are in use, and it seems to be useful compare their properties.

Finally, the plan of paper includes: in section II and III the local sampling and the technique of nearest neighbours method was briefly discussed. In section IV and V the procedure of joint non-parametric linearisation was explained and families of linear lines with fixed and random parameters were briefly discussed. At last, in VI section the known procedures for estimate a measurement error were examined with Monte Carlo modelling.

II. LOCAL SAMPLING WITH k-NN TECHNIQUE

In applied statistics a local sampling usually constitutes the initial step to further non-parametric analysis like, for example, a local regression. A local sample, drawn from some whole sample, contains all individuals, which are sufficiently similar each to other or to assumed pattern. Similarity between individuals is often defined by notion of distance in space of the observed variables, but it is not essential (Dette and Gefeller 1995). Nevertheless, there are a lot of approaches to local sampling. First of all we should distinguish partitioning on the non-overlapping clusters and other methods. The simple strategy for non-overlapping clustering is to categorise all observed variables separately and then form cells as combinations of the categories. A drawback of this strategy is that number of combinations gets large even for moderate (e.g. two or three) numbers of categories. It usually leads to loss of some data, because many cells incidences will be to small for assumed further analysis. The more sophisticated procedures can divide given set of individuals into approximately equal clusters. Nevertheless, any non-overlapping partitioning in practice meet usually with contradiction between accessible number of all data, number of considered variables, and postulated number of clusters and number of data in each cluster. Thus, if further analysis is foreseen, then usually only overlapping partitioning has practical meaning. In this two approaches can be distinguish. First one admits that each individual can belong simultaneously to each singled out local sample or, in other words, that each local sample can includes all considered set of individuals. Consequently, the individuals belong to separate local samples not at all, but with some weight, associated with notion of kernel function or with member function (Keming and Jones 1998). The second approach admits, that each individual can on the whole belong simultaneously to some singled out local samples, at least some individuals can belong to each local sample, but each local sample contain only some part of all considered sample. It can be interpreted in the terms of the first approach in such way, that some individuals belong to considered local sample with weight equal 1 and all remain individuals with weight equal 0.

Suppose a given data set $Y_1^T, Y_2^T, ..., Y_J^T, Y_{J+1}^T, Y_{J+2}^T, ..., Y_N^T$, where $Y_1^T, Y_2^T, ..., Y_J^T$ are assumed seeds of local samples, which includes K points mostly similar to its seed, from given sample of points $Y_1^T, Y_2^T, ..., Y_J^T$, and maybe some additional ones, $Y_{J+1}^T, Y_{J+2}^T, ..., Y_N^T$. A member shares w_{ij}, of i-th point into j-th local sample, $i = 1, 2, ..., N$, $j = 1, 2, ..., J$, are equal 1 or 0. It means, that any i-th point can all belong to $L_i > 1$ local samples simultaneously. A sum of member shares w_{ij}, for each j-th local sample constitutes a number K_j

of members inside this local sample. It is assumed, that any separate analysis, which concerns relations exclusively inside single local sample, is founded on this number K_j of data. Numbers K_j can be used to estimate a significance of result in each single local sample treated separately. Nevertheless, if any analysis concerns all local samples simultaneously, then each local sample represents not K_j number of data, but only $U_j = \Sigma\ (w_{ij} / L_i)$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, J$, number of data. In practice usually each or almost each local sample includes the same number K of members. Then, if the differences between L_i for $i = 1, \ldots, N$, are neglected, the numbers of data could be assumed equal $U_j = U =$ idem; $j = 1, \ldots, J$, with $U = \min(1, N / J)$, because the method did not limit number J of local samples, and each local sample was treated as a source of single data. Moreover, if any analysis includes comparisons between the local samples and the Bonferroni adjustment should be taken into account, then number of independent samples could be estimated as $N / K << J$.

III. NEAREST NEIGHBOURS (k-NN) TECHNIQUE

The *k-NN* technique assumes, that similarity between individuals is defined by Euclidean distances between points in the space of the L observed variables Y^T, where $Y^T = (Y_1, \ldots, Y_L)$. Usually it is supplement with hypothesis, that given data set was drawn from uniform distribution in a proper parallelepiped in space of Y^T (Ripley 1979). If this hypothesis is true, the natural assumption is, that each local sample should represent the same probability, what be expressed in the demand, that each local sample should include the same number K of members, or in the demand, that each local sample should get the same volume in the parallelepiped. Both above demands lead to so named edge effect (Doguwa and Upton 1988): the ideal local samples of the same probability should be represent by greater spheres near to edge than in the middle of a parallelepiped. Nevertheless, in practice a hypothesis of multivariate uniform distribution is often replaced with non-equivalent set of L separate requirements, that each variable $Y_1, \ldots, Y_L$ must be previously transformed to univariate uniform distribution on the interval $(0, G_l)$, where G_l exemplify importance of a l-th variable Y_l, $l = 1, \ldots, L$. In such situation it is possible, that edge effect appears inside the parallelepiped $(0, G_l)$, $l = 1, \ldots, L$, too.

In practical applications a minimal number K of data in local sample follows from purpose of analyse. From other hand, a maximal distance inside local sample cannot oppose to presumption, that differences between members can be neglected. If any such information is not available, than it seems be reasonable put $G_l = 1$ for all $l = 1, \ldots, L$; and start with $7 \div 12 < K < 30 \div 40$ and maximal distance between seed and member of local sample $D < 0{,}20 \div 0{,}25$. In proposed procedure a role of a seed of local sample previously plays each data from a given data set $Y_1^T, Y_2^T, ..., Y_J^T, Y_{J+1}^T, Y_{J+2}^T, ..., Y_N^T$. For some assumed values of K the numbers J of local samples with assumed D should be computed, and on this base the final decision on K, D and J can be taken. All local samples with greater D should be excluded from further analyses.

IV. LINEARIZATION BY MONOTONIC TRANSFORMATION OF AN INDEPENDEND VARIABLE

Suppose strong ordering $X_1^{'} < X_2^{'} < ... < X_N^{'}$. A definition on monotonic transformation $X_i^{'} \rightarrow X_i$ implies, that $(X_i^{'} < X_k^{'}) \rightarrow (X_i < X_k)$ for all i, k = 1, ..., N. Function (Z_i, X_i) is a monotonic (and increasing) function if $(X_i < X_k) \rightarrow (Z_i < Z_k)$ for all i, $k = 1, ..., N$; it is a monotonic (and decreasing) function if $(X_i < X_k) \rightarrow (Z_i > Z_k)$ for all i, $k = 1, ..., N$. If transformation $X_i^{'} \rightarrow X_i$ is a monotonic one, then a monotonic function $(Z_i, X_i^{'})$ persists to be monotonic function after this transformation, and a non-monotonic function $(Z_i, X_i^{'})$ persists to be monotonic function after this transformation. Linear function $Z_i = b_0 + b_1X_i$, where b_0 and $b_1 \neq 0$ are fixed constants, $i = 1, ..., N$, is a monotonic function. Consequently, only any monotonic function $(Z_i, X_i^{'})$ can be exactly transformed into any linear function with the monotonic transformation $X_i^{'} \rightarrow X_i = (Z_i - b_0) / b_1$ into a linear function $Z_i = b_0 + b_1X_i$, where b_0 and $b_1 \neq 0$ are freely chosen constants.

Consider a non-monotonic function $(Z_i, X_i^{'})$. Let us look for minimal number M_i that $Z_i < \min(Z_{i+M+1}, Z_{i+M+2}, ..., Z_N)$, for each Z_i, $i = 1, ..., N - 1$, and if $M_i > 0$ let us change Z_{i+m} with $Z_{i+m}^{'} = Z_i^{\wedge} + m^{*}\mathbf{0}$, where: $Z_i^{\wedge} = (Z_{i+1}, Z_{i+2}, ..., Z_M) / m$; $m = 1, ..., M$; and $\mathbf{0}$ is a neglected small number. After above procedure considered function (Z, X) get monotonicity and it can be exactly linearesed. The differences $\varepsilon_{i+m} = Z_{i+m} - Z_{i+m}^{'}$ can be treated as the errors of linearisation. They don't depend on freely chosen parameters of linear function b_0 and $b_1 \neq 0$. They don't depend on any values of variable X, and they don't depend on any values of variable Z exept $Z_{i+1}, Z_{i+2}, ..., Z_M$. It can be proved that for any chosen constants b_0 and $b_1 \neq 0$ a value $X_i\wedge = (Z_i^{\wedge} - b_0) / b_1$ leads to minimal value of $\Sigma_i(\varepsilon^2) = \varepsilon^2_{i+1} + \varepsilon^2_{i+2} + ... + \varepsilon^2_{i+M}$. Consequently, each disorder of monotonicity can be optimal linearised independently from others disorders under general criterion $\Sigma\Sigma(\varepsilon^2)$. Moreover, the monotonic linearisation can be formulated as task: find $(X_1, X_2, \ldots, X_N)$ which leads to minimal value of $\Sigma\Sigma_i(\varepsilon^2)$, under restrictions $X_1 < X_2 < \ldots < X_N$.

Suppose now, that the known values of Z_i were charged with random errors e_i, $i = 1, ..., N$. Probability $P(Z_j < Z_i)$ depends on the distributions of errors e_i and e_j. Suppose that for $j < i < N$ probability $P(Z_j < Z_i) \geq P$; $i, j = 1, ..., N$. An event $\varepsilon_i = 0$ take place, if $(M_1 < i - 1\ )$ and $(M_2 < i - 2\ )$ and ... and $(M_{i-1} < 1)$ and $(M_i < 1\ )$. Consequently, probability $P(\varepsilon_i = 0) \geq P^{(N-i+1)*i-1}$, and probability $P(\Sigma\varepsilon^2 = 0) = P((\varepsilon_1 = 0)$ and $(\varepsilon_2 = 0)$ and ... $(\varepsilon_{N-1} = 0)) \geq P^{(N-1)*N/2}$. For example, for $N = 8$ and $P = 0{,}975$; $P(\Sigma\varepsilon^2 = 0) \geq 0{,}49$; for $N = 8$ and $P = 0{,}998$; $P(\Sigma\varepsilon^2 = 0) \geq 0{,}95$. It should be noted, that $P(\varepsilon_i = 0) \geq 0$ and $P(\Sigma\varepsilon^2 = 0) \geq 0$ for continuous (e.g. normal) distributions of errors e_i, $i = 1, ..., N$. If event $(\Sigma\varepsilon^2 > 0)$ take place, and errors e_i are distributed normally, with assumed the same (unknown) standard deviation, then the distribution of $\Sigma\varepsilon^2/V_\varepsilon$ can be approximated by χ^2 distribution with degree of freedom $df = \Sigma M_i$, where variation V_ε of ε is estimated only for groups of M_i data with $\Sigma\varepsilon^2 > 0$.

V. JOINT LINEARISATION BY MONOTONIC TRANSFORMATION OF AN INDEPENDEND VARIABLE

Consider now some given samples (Z_{1i}, X'_{1i}), (Z_{2i}, X'_{2i}), …, (Z_{Ji}, X'_{Ji}), where in each sample $i = 1, \ldots, N_j$. Suppose that for each sample (Z_{ji}, X'_{ji}), $i = 1, \ldots, N_j$, it could be find at least a single sample (Z_{ki}, X'_{ki}), $i = 1, \ldots, N_k$, that maximum $(X_{j1}, X_{k1}) <$ minimum (X_{j1}, X_{k1}). Note that for any pair of samples with maximum$(X_{j1}, X_{k1}) >$ minimum(X_{j1}, X_{k1}) the task of simultaneous linearisation divides into two separate tasks and joint linearisation don't occurs.

Let all given values of X' were ordered $X'_1 < X'_2 < \ldots < X'_N$, where $N = N_1 + N_2 + \ldots + N_J$. Then task of joint linearisation can be formulated as follow: find $(X_1, X_2, \ldots, X_N)$ which leads to minimum(maximum$(\Sigma\Sigma_i(\varepsilon^2)_1, \Sigma\Sigma_i(\varepsilon^2)_2, \ldots, \Sigma\Sigma_i(\varepsilon^2)_J)$) under restrictions $X_1 < X_2 < \ldots < X_N$. This task can be solved with the iterative procedure: first for each given sample the parameters b_{0j} and b_{1j} of regression $Z_j^\wedge = b_{0j} + b_{1j}{*}X$, $j = 1, \ldots, J$, should be computed with the last square errors criterion. Then criterion $C =$ maximum$(\Sigma\Sigma_i(\varepsilon^2)_1, \Sigma\Sigma_i(\varepsilon^2)_2, \ldots, \Sigma\Sigma_i(\varepsilon^2)_J)$ should be computed. In each following step of procedure it is tested, whether exist transformation $X \rightarrow X$ which gives a smaller C then in a previous step, or which gives the same C then in a previous step but it gives a smaller sum $\Sigma(\Sigma\Sigma_i(\varepsilon^2)_1, \Sigma\Sigma_i(\varepsilon^2)_2, \ldots, \Sigma\Sigma_i(\varepsilon^2)_J)$. If such transformation exists, and it satisfies restrictions on ordering of X's, then it should be performed, if not – it leads to the end of procedure. In practical realisation an initial restriction $X_i < X_{i+1}$ should be formulated as $X_{i+1} - X_i > h$, $i = 1, N - 1$, where small constant h should be chosen accordingly to differences between X's, e.g. as about 0,000001 of mean difference. The procedure should be stopped, if a improvement of criterion C is less then about 0,1% of its previous value. With such constants procedure stopped after about (1 ÷ 3)*J iterations (Górkiewicz and Kawalec 1999).

It should be marked, that the joint linearisation distinctly differs from the separate linearisation. Let us explain it on example of two monotonic samples (Z_{1i}, X'_{1i}) and (Z_{2i}, X'_{2i}), where $X'_{1i} = X'_{2i}$; $i = 1, \ldots, N$. The separate linearisation always leads to $\Sigma\Sigma_i(\varepsilon^2)_1 = \Sigma\Sigma_i(\varepsilon^2)_2 = 0$, because for any monotonic function $\Sigma\Sigma_i(\varepsilon^2) = 0$. The joint linearisation leads to $\Sigma\Sigma_i(\varepsilon^2)_1 = \Sigma\Sigma_i(\varepsilon^2)_2 = 0$ only if there exist constants a_0 and $a_1 \neq 0$ that $Z_{1i} = a_0 + a_1 {*} Z_{21}$; $i = 1, \ldots, N$. Thus, in joint linearisation of J samples in practice a minimal value of criterion minimum(maximum($\Sigma\Sigma_i(\varepsilon^2)_1$, $\Sigma\Sigma_i(\varepsilon^2)_2$, ..., $\Sigma\Sigma_i(\varepsilon^2)_J$)) occurs with all $\Sigma\Sigma_i(\varepsilon^2)_j > 0$, $j = 1, \ldots, J$.

The proposed procedure for joint linearisation always leads to minimal value of assumed criterion. Nevertheless, the minimal value can be find to much, and a few mostly troubled samples should be excluded from joint linearisation, or given set of samples should be divided into two or three parts with appropriate values of criterion.

Some statistical approaches suit to further analyse of the approved results of joint linearisation. The simplest approach assumes that a given initial data set (Z, X, Y^T) was transformed into a new data sample (b_{0j}, b_{1j}, Y_j^T), $j = 1, \ldots, J$; where Y_j^T are assumed representation of j-th local sample; b_{0j} and b_{1j} are assumed as random numbers drawn from normal distributions $N(0, \sigma_{0j})$ and $N(0, \sigma_{1j})$, where σ_{0j} and σ_{1j} are estimated as the sample standard deviations SD_{0j} and SD_{1j}. Within this approach the regressions b_0^ = $f(Y^T)$ and b_1^ = $f(Y^T)$ can be analysed. Nevertheless, hypothesis b_{0j}, = idem or b_{1j} = idem or $b_{0j} + X_0 {*} b_{1j}$ = idem can be tested, where constant X_0 represents a common cut point of all J lines. It is known, that under assumption b_{1j} = idem a statistics CHI = $\Sigma(b_{1j} / SD_{1j})^2 - (\Sigma(b_{1j} / (SD_{1j})^2)^2 / \Sigma\, 1/(SD_{1j})^2$, $j = 1, \ldots, J$; is a chi-square variable with $df = J - 1$ degree of freedom. A combined estimate of slope is $b_1^\wedge = \Sigma(b_{1j} / (SD_{1j})^2 / \Sigma\, 1/ (SD_{1j})^2$. If obtained value of CHI is greater then critical value of chi-square test, then hypothesis b_{1j} = idem don't valid. In such situation one can exclude a few most confusing b_{1j}'s and try the above analyse once again (Omar et al., 1999).

The more sophisticated approach takes into account all resulting local samples, and consider them as separate random samples. Within this approach the assays of regressions lines $Z_j^\wedge = a_j + b_j {*} X$ are estimated and tested apart from parameters b_{0j} and b_{1j} obtained with procedure of joint linearisation. Here, when the hypothesis b_j = idem of equal slope parameter is rejected, or hypothesis $a_j + X_0{*}b_j$ = idem is rejected, then the Johnson-Neyman technique can be used to to determine a region of the independent variable for which no significant differences of outcome Z can be detected (Schwenke 1990). Nevertheless, instead of partitioning on the set of local samples, in the considering models hypothesis of fixed parameters a_j and b_j can be changed with hypothesis of random parameters (Hildrecht and Houck 1968, Longford 1995). Methods for statistical analyse of the linear assays were examined by many researchers (Hanusz 2000, Heckman and Zamar 2000, Jensen 1989, Srivastava et al. 1980) and they were implemented in some known statistical packages. It those ways of analyse don't result successfully, the previous approach must be applied. In a successful case for each considered local sample its own representation should be chosen, and regression $a^\wedge = f(Y^T)$ and $b^\wedge = f(Y^T)$ can be analysed.

VI. ESTIMATION OF STANDARD DEVIATION BY REPEATED MEASUREMENTS

The method proposed in the above sections can be applied to estimation of models with measured value as independent variable X and standard deviation $SDE(X)$ of measurements error as outcome Z. Of course, a standard deviation cannot be measured directly. So, in the paper the following procedure was examined. For each considered i-th individual, $i = 1, \ldots, N$; a set of repeated measurements $X(i, r)$ was achieved, $r = 1, \ldots, R_i$. If it can be assumed, that the true measured value X and the standard deviation of measurements error $\sigma(X)$ were constant over time of repeated measurements, then a sample mean $\Sigma\, X(i, r)$, $r = 1, \ldots, R_i$; estimates a true value of X for i-th individual, and sample standard deviation of measurements estimates $X(i, r)$, $r = 1, \ldots, R_i$; estimates a true value of $\sigma(X)$ for i-th individual.

In order to compare the three known estimates of standard deviation a sets of 1000 individuals was modelled. For each individual a set of R values of X was generated from the normal distribution $N(0, 2)$ for $R = 2, 3, \ldots, 11$. A sample standard deviation for each set of R measurements X was estimated as a single number with the classical formula $SD_1(X) = \sqrt{}\ \Sigma((X_r - X^\wedge) * (X_r - X^\wedge) / (R - 1))$ and with the formula $SD_2(X) = \Sigma\ |X_r - X^\wedge| / (R - 1)$, and as the sample of R numbers $SD_r(X) = R * |X_r - X^\wedge| / (R - 1)$, where a mean value $X^\wedge = \Sigma X_r / R$, $r = 1, 2, \ldots, R$. It was confirmed that for $2 \leq R \leq 11$ all considered estimates are unbiased, it means they practically don't differ from assumed in modelling value

$\sigma(X)$. Moreover, on the base on 1000 data for each $R = 2, 3, \ldots, 11$ the standard deviation $SD(SD_1)$, $SD(SD_2)$ and $SD(SD_r)$ were estimated. It was stated, that for $R = 2$: $SD(SD_2) = SD(SD_r) \cong 0.85*\sigma(X)$ and $SD(SD_1) \cong 0.6*\sigma(X)$. For $R = 3$: $SD(SD_r) \cong 0.75*\sigma(X)$ and $SD(SD_1) \cong SD(SD_2) \cong \sigma(X)$. For $4 \leq R \leq 11$: $SD(SD_r) \cong 0.7*\sigma(X) \cong$ idem, and practically $SD(SD_1) = SD(SD_2)$. Besides, in each separate sample of R data a sample standard deviation $SD_R(X)$ and sample standard deviation $SD_R(SD_r)$ was computed and for $4 \leq R \leq 11$ significant regression $SD_R(SD_r) \cong 0.6 \div 0.7\ SD_R(X)$ was confirmed. Thus, estimate SD_r provides an self-correcting property: samples with greater random sample deviation $SD_R(X)$ have the greater sample deviation $SD_R(SD_r)$ and they weekly exert on regression between measured value X and estimated standard deviation of measurement error then other samples with smaller random sample deviation $SD_R(X)$. It gives reason for conclusion, that in considered task the use of estimate SD_r can be recommended. Nevertheless, the another problem arises, because if the independent variable is measured as mean of some repeated measures with not neglected standard deviation, than conventional parametric regression methods are no valid (Caroll and al. 1999). Thus, it should be recommended to choose number R of repeated measures under criterion $SD_R(X) << SD(SD_1) \cong SD(SD_2)$.

REFERENCES

Bland J. M., Altman D. G. (1986), *Statistical Methods for Assessing Agreement Between two Methods of Clinical Measurement*, "Lancet", **I**, 307–310.

Carrol R. J., Maca J. D., Ruppert D. (1999), *Nonparametric Regression in the Presence of Measurement Error*,"Biometrika", **86**, 3, 541–554.

Crowder M. J, Hand D. J. (1990), *Analysis of Repeated Measures*. Chapman and Hall, London.

Davidian M., Giltinan D. M. (1995), *Nonlinear Models for Repeated Measures*. Chapman and Hall, London.

Dette, H., Gefeller, O. (1995), *The Impact of Different Definitions of Nearest Neighbour Distances for Censored Data on the Nearest Neighbour Kernel Estimators of the Hazard Rate*. "Journal of Nonparametric Statistics", **4**, 271–282.

Doguwa S. I., Upton G. J. G. (1988), *On Edge Correction for the Point-Event Analogue of the Clark-Evans Statistic*, "Biometrical Journal", **30**, 8, 957–963.

Doguwa S. I., Upton G. J. G. (1990), *On the Estimation of the Nearest-Neighbour Distribution G(t) for Point Processes*, "Biometrical J", **32**, 7, 863–876.

Domański Cz. (1990), *Testy statystyczne*, PWE, Warszawa.

Domański Cz., Pruska K. (2000), *On Unemployment Investigation in Small Areas*, "Acta Universitas Lodziensis", Folia Oeconomica **152**, 99–115.

Górkiewicz M., Kawalec E. (2000), *Estimation of non-Cox Proportional Hazard by k-Nearest Neighbours Sampling and Transformation of Local Hazard Estimates. Proc of Statistics and Clinical Practice*, 82–85, Warszawa.

Hanusz Z. (2000), *Relative Potency for the Multivariate Contaminated Normal Responses*, "Acta Universitas Lodziensis", Folia Oeconomica 152, 127–139.

Heckman N. E., Zamar R. H. (2000), *Comparing the Shapes of Regression Functions*, "Biometrika", **87**, 1, 135–144.

Hildrecht C., Houck J. P. (1968), *Some Estimators for a Linear Model with Random Coefficients*, "Journal of the American Statistical Association", **63**, 584–595.

Jensen D. R. (1989), *Joint Confidence Sets in Multiple Dilution Assays*, "Biometrical Journal", **31**, 7, 841–853.

Jianquing F., Sheng-Kuei L. (1998), *Test of Significance when Data are Curves*, "Journal of the American Statistical Association", **93**, 443, 1007–1021.

Keming Y., Jones M. C. (1998), *Local Linear Quantile Regression*, "Journal of the American Statistical Association", **93**, 441, 228–237.

Korzeniewski J. (2000), *Sample Breakdown Point of the Wilcoxon and Sign Tests for Location*. "Acta Universitas Lodziensis", Folia Oeconomica **152**, 93–98.

Longford N. (1995), *Random Coefficient Models*, Oxford Science Publications. Oxford.

Omar R. Z., Wright E. M., Turner R. M., Thompson S. G. (1999), *Analysing Repeated Measurements Data: A Practical Comparisons of Methods*, "Statistics in Medicine", **18**, 1587–1603.

Rao C. R. (1975), *Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems*, "Biometrics", **31**, 545–554.

Ripley B. D. (1979), *Tests of Randomness for Spatial Point Patterns*, "J. Roy. Statist. Soc". Series B, **41**, 368–374.

Schwenke J. R. (1990), *On the Equivalence of the Johnson-Neyman Technique and Fieller's Theorem*, "Biometrical Journal", **32**, 4, 441–447.

Srivastava V. K., Bhattacharya B. N., Kumar K. (1980), *Improved Estimation of Potency in Slope Ratio Assays*, "Biometrical Journal", **22**, 1, 61–66.

Maciej Górkiewicz

ESTYMOWANIE BŁĘDU POMIAROWEGO Z ZASTOSOWANIEM ŁĄCZNEJ NIEPARAMETRYCZNEJ LINEARYZACJI PRÓB LOKALNYCH

Praca prezentuje zastosowanie techniki najbliższych sąsiadów w celu przekształcenia zbioru N danych postaci (Z, X, Y^T) w zbiór $J \approx N$ prób lokalnych (Z, X), przy ograniczeniach dotyczących minimalnej liczby danych K oraz różnic wartości Y^T w każdej próbie lokalnej, gdzie Z pełni rolę zmiennej zależnej, X – zmiennej niezależnej, a $Y^T = (Y_1, ..., Y_L)$ jest L-wymiarową zmienną dodatkową. Następnie proponuje się procedurę nieparametrycznej łącznej linearyzacji zbioru prób lokalnych. Obie procedury proponuje się stosować do oceny dokładności metod pomiarowych, z odchyleniem standardowym błędu pomiarów jako zmienną Z i wielkością mierzoną jako zmienną X. Proponowane podejście może być użyteczne w innych zastosowaniach, kiedy zamiast modelu regresji wielowymiarowej estymuje się rodzinę zależności regresyjnych.