Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 304 · 2016

Informatyka i Ekonometria 7

Grzegorz Sitek

Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
Katedra Statystyki
grzegorz.sitek@ue.katowice.pl

ESTIMATION OF REGRESSION PARAMETERS OF TWO DIMENSIONAL PROBABILITY DISTRIBUTION MIXTURES

Summary: We use two methods of estimation parameters in a mixture regression: maximum likelihood (MLE) and the least squares method for an implicit interdependence. The most popular method for maximum likelihood esti-mation of the parameter vector is the EM algorithm. The least squares method for an implicit interdependence is based solving systems of nonlinear equations. Most frequently used method in the estimation of parameters mixtures regression is the method of maximum likelihood. The article presents the possibility of using a different the least squares method for an implicit interdependence and compare it with the maximum likelihood method. We compare accuracy of two methods of estimation by simulation using bias: root mean square error and bootstrapping standard errors of estimation.

Keywords: mixture regression model, EM algorithm, least squares method for an implicit interdependence.

Introduction

The use of mixtures of regressions falls into two primary categories. The first involves estimating a set of regression coefficients for all observations coming from a possibly unknown number of heterogeneous classes. A second use for mixtures of regressions is in outlier detection or robust regression estimation. This occurs under the assumption that one regression plane can adequately model the data, but there is an apparent class heterogeneity because of large variances attributed to some observations which are considered outliers.

Mixtures of regressions have been extensively studied in the econometrics literature and were first introduced by Quandt [1972] as the switching regimes (or switching regressions) problem. A switching regimes system is often compared to structural change in a system [Quandt, Ramsey, 1978]. The problem of parameters estimation of switching regression was presented by Pruska [1992]. A structural change assumes the system depends deterministically on some observable variables (such as t in the switching point data), but switching regimes implies one is unaware of what causes the switch between regimes.

Mixture models have been widely used in econometrics and social science, and the theories for mixture models have been well studied [Lindsay, 1995]. The classical Gaussian mixture case has been extensively studied [McLachlan, Peel, 2000]. The robust mixture regression procedure based on the skew t distribution was presented by Dogru and Arslan [2015].

The main aim of this work is comparison by simulation the accuracy of the estimators obtained by the maximum likelihood and the least squares method for an implicit interdependence.

1. The least squares method for an implicit interdependence

Two lines, none of which is parallel to the axis of the system, are described by the equation [Antoniewicz, 1988]:

$$(y - ax - b)(y - cx - d) = 0 \tag{1}$$

Based on the available data, we wish to estimate the parameters a, b, c and d. This research is equivalent to finding the straight lines that gives the best fit (representation) of the points in the scatter plot of the response versus the predictor variable (see: Figure 1). We estimate the parameters using the popular least squares method which gives the lines that minimizes the sum of squares of the vertical distances from each point to the lines. The vertical distances represent the errors in the response variable.

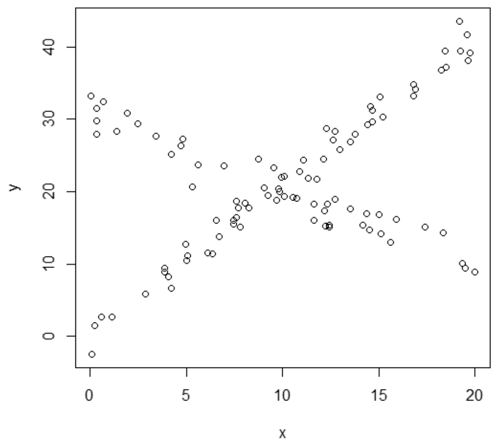

Figure 1. Scatter plot of *y* versus *x*

The sum of squares of these distances can then be written as:

$$S(a,b,c,d) = \sum_{i=1}^{n} [(y_i - a \cdot x_i - b)(y_i - c \cdot x_i - d)]^2 \tag{2}$$

The values of $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ that minimize S(a,b,c,d) are given by solving systems of nonlinear equations (4). To shorten the notation we use the following notation:

$$[x^r] = \sum_{i=1}^{n} x^r \tag{3}$$

$$\begin{cases} ac^2[x^4] + ad^2[x^2] + bc^2[x^3] + bd^2[x] + 2bcd[x^2] + 2acd[x^3] - c^2[x^3 y] \\ - d^2[xy] - 2ac[x^3 y] - 2bc[x^2 y] - 2cd[x^2 y] - 2bd[xy] - 2ad[x^2 y] \\ + a[x^2 y^2] + 2c[x^2 y^2] + 2d[xy^2] + b[xy^2] - [xy^2] = 0 \\ bc^2[x^2] + bd^2 + ac^2[x^3] + ad^2[x] + 2acd[x^2] + 2bcd[x] - 2ad[xy] - 2bc[xy] \\ - 2cd[xy] - 2bd[y] - d^2[y] - c^2[x^2 y] - 2ac[x^2 y] + b[y^2] \\ + 2d[y^2] + 2c[xy^2] + a[xy^2] - [xy^3] = 0 \\ a^2 c[x^4] + cb^2[x^2] + da^2[x^3] + db^2[x] + 2abd[x^2] + 2abc[x^3] - 2ac[x^3 y] \\ - d^2[xy] - 2ac[x^3 y] - 2ab[x^2 y] - 2ad[x^2 y] - b^2[xy] - 2bd[xy] - a^2[x^3 y] \\ - 2bc[x^2 y] + 2a[x^2 y^2] + c[x^2 y^2] + 2b[xy^2] + d[xy^3] - [xy^3] = 0 \\ da^2[x^4] + b^2 d + a^2 c[x^3] + cb^2[x] + 2acb[x^2] + 2bad[x] - 2ac[x^2 y] \\ - 2ab[xy] - 2ad[xy] - 2bc[xy] - 2bd[y] - a^2[x^2 y] - b^2[y] + 2a[xy^2] \\ + 2b[y^2] + c[xy^2] + d[y^2] - [y^3] = 0 \end{cases} \tag{4}$$

The least squares regression lines are given by:

$$\hat{y}_1 = \hat{a} \cdot x + \hat{b}, \quad \hat{y}_2 = \hat{c} \cdot x + \hat{d} \tag{5}$$

2. Mixtures of linear regressions

In search of a straightforward way to build a regression procedure, we return to the standard definition of the regression function m(x):

$$m(x) = E(Y \mid X = x) = \int yf(y \mid x)dy = \frac{\int yf(x,y)dy}{\int f(x,y)dy} \tag{6}$$

It is well-known that when f(x,y) is Gaussian, the conditional pdf f(y|x) is Gaussian and the regression function m(x) is linear. A natural extension of the single Gaussian pdf for f(x,y) is to model f(x,y) as a K-component Gaussian mixture:

$$f(x,y) = \sum_{j=1}^{K} p_j \phi(x, y, \mu_j, \Sigma_j) \tag{7}$$

where:

ϕ – the bivariate Gaussian density, $\sum_{j=1}^{K} p_j = 1$;

$\mu_j = \begin{bmatrix} \mu_{jX} \\ \mu_{jY} \end{bmatrix}$ – the expected values of j-th component of the mixture;

$\Sigma_j = \begin{bmatrix} \Sigma_{jX} & \Sigma_{jXY} \\ \Sigma_{jYX} & \Sigma_{jY} \end{bmatrix}$ – covariance matrix of j-th component of the mixture.

The resulting regression function m(x) of the pdf in (7) is a combination of linear functions $m_j(x)$. That is:

$$m(x) = \sum_{j=1}^{K} w_j(x) m_j(x) \tag{8}$$

The conditional variance function is:

$$v(x) = Var[Y \mid X = x] = \sum_{j=1}^{K} w_j(x)(m_j(x) + \sigma_j) - (\sum_{j=1}^{K} w_j(x) m_j(x))^2 \tag{9}$$

Let K=2. The bivariate distribution (X,Y) is a mixture bivariate distributions (X_1,Y_1) and (X_2,Y_2):

$$f(x,y) = pf_1(x,y) + (1-p) f_2(x,y) \tag{10}$$

Thus, let:

$$h(x)=\int f(x,y)dy=p\int f_1(x,y)dy+(1-p)\int f_2(x,y)dy=ph_1(x)+(1-p)h_2(x) \quad (11)$$

$$f(y\,|\,x)=\frac{pf_1(x,y)+(1-p)f_2(x,y)}{ph_1(x)+(1-p)h_2(x)}=f_1(y\,|\,x)w_1(x)+f_2(y\,|\,x)w_2(x) \quad (12)$$

where:

$$w_1(x)=\frac{ph_1(x)}{h(x)},\, w_2(x)=\frac{(1-p)h_2(x)}{h(x)},$$

if: $h_1(x)=h_2(x),\quad h(x)=h_1(x)=h_2(x)$.

Equality $h_1(x)=h_2(x)$ is true, if the distributions (X_1,Y_1) and (X_2,Y_2) differ only correlation.

$$f(y\,|\,x)=pf_1(y\,|\,x)+(1-p)f_2(y\,|\,x) \quad (13)$$

$$E(Y\,|\,x)=pE(Y_1\,|\,x)+(1-p)E(Y_2\,|\,x) \quad (14)$$

In particular:

$$E(Y\,|\,x)=p(a_1x+b_1)+(1-p)(a_2x+b_2) \quad (15)$$

Suppose we have n independent univariate observations, y_1, y_2,…, x_n, each with a corresponding vector of predictors, x_1,…, x_n with $x_i = (x_{i,1}, x_{i,2})^T$ for i = 1,…, n. We often set $x_{i,1} = 1$ to allow for an intercept term. Let $y = (y_1,…, y_n)^T$ and let $\underline{X}$ be the matrix consisting of the predictor vectors. Suppose further that each observation (y_i; x_i) belongs to one of two classes. Conditional on membership in the j-th component, the relationship between y_i and x_i is the normal regression model:

$$y_i=x_i^T\beta_j+\varepsilon_i \quad (16)$$

where $\varepsilon_i \sim N(0,\,\sigma_j^2)$ and β_j,σ_j^2 are the two-dimensional vector of regression coefficients and the error variance for component *j*, respectively. Accounting for the mixture structure, the conditional density of $y_i|x_i$ is:

$$g_\theta(y_i\,|x_i)=\sum_{j=1}^{2}p_j\phi(y_i\,|x_i^T\,\beta_j,\sigma_j^2) \quad (17)$$

where $\phi(y_i\,|x_i^T\,\beta_j,\sigma_j^2)$ is the normal density with mean $x^T\beta_j$ and variance σ_j^2.

2.1. EM algorithm for mixtures of regressions

The landmark article that introduces the EM algorithm is [Dempster, Laird, Rubin, 1977]. The EM algorithm is used to find locally maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. There are several reasons for its popularity. The main advantages of the EM algorithm are its simplicity and ease of implementation. In addition, the EM algorithm always converges monotonically. On the other hand, EM algorithm also suffers some disadvantages. The main disadvantage of the EM algorithm is its slow linear convergence. Also, the EM algorithm results depend on the initial value. Hence, the key for the success of EM algorithm is a good initial guess. EM algorithm is the standard algorithm for computing the MLE of the Gaussian mixture models. Given the number of component K, the EM algorithm for Gaussian Mixtures is straightforward.

A standard EM algorithm may be used to find a local maximum of the likelihood surface. De Veaux [1989] describes EM algorithms for mixtures of regressions in more detail, including proposing a method for choosing a starting point in the parameter space.

E-step: Calculate the posterior probabilities of component inclusion:

$$p_{ij}^{(t)} \overset{def}{=} \frac{p_j^{(t)} \phi(y_i \,|\, x_i^T \beta_j, \sigma_j^2)}{\sum_{l=1}^{2} p_l^{(t)} \phi(y_i \,|\, x_i^T \beta_l, \sigma_l^2)} \tag{18}$$

for $i = 1,2,\ldots, n$ and $j = 1,2$.

Numerically, it can be dangerous to implement equation (18) exactly as written due to the possibility of the indeterminant form 0/0. Thus, many of the routines in mixtools [Benaglia, Chauveau, Hunter, Young, 2009] actually use the equivalent expression:

$$p_{ij}^{(t)} = \left[1 + \sum_{l \neq j} \frac{p_l^{(t)} \phi(y_i \,|\, x_i^T \beta_l, \sigma_l^2)}{p_j^{(t)} \phi(y_i \,|\, x_i^T \beta_j, \sigma_j^2)} \right]^{-1} \tag{19}$$

M-step for p:

$$p_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} p_{ij}^{(t)} \quad \text{for } j = 1,2 \tag{20}$$

Letting:

$$W_j^{(t)} = diag(p_{1j}^{(t)}, \ldots, p_{nj}^{(t)})$$

the additional M-step updates to the β and σ parameters are given by:

$$\beta_j^{(t+1)} = (\underline{X}^T W_j^{(t)} \underline{X})^{-1} \underline{X}^T W_j^{(t)} y \tag{21}$$

and:

$$\sigma_j^{2(t+1)} = \frac{\left\| W_j^{\frac{1}{2}(t)} (y - \underline{X}^T \beta_j^{(t+1)}) \right\|^2}{tr(W_j^{(t)})} \tag{22}$$

where $\|A\|^2 = A^T A$ and tr(A) means the trace of the matrix A. Notice that equation (21) is a weighted least squares (WLS) estimate of β_j and equation (22) resembles the variance estimate used in WLS. Allowing each component to have its own error variance σ_j^2 results in the likelihood surface having no maximizer, since the likelihood may be driven to infinity if one component gives a regression surface passing through one or more points exactly and the variance for that component is allowed to go to zero. A similar phenomenon is well-known in the finite mixture of normal model where the component variances are allowed to be distinct [McLachlan, Peel, 2000]. However, in practice we observe this behavior infrequently and the mixtools functions automatically force their EM algorithms to restart at randomly chosen parameter values when it occurs. The function regmixEM implements the EM algorithm for mixtures of regressions in mixtools.

3. Comparison of estimation accuracy

This chapter presents the results of simulations. The simulations performed for three different values of p: 0,5; 0,7; 0,8. Then, based on randomly generated data, we estimate the parameters a, b, c, d using the least squares method for an implicit interdependence and MLE method. We compare both methods using bias and root mean square error (RMSE) of estimator. We use the following formulas:

$$\hat{\theta} = \frac{\sum_{i=1}^{N} \hat{\theta}_i}{N},\ RMSE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{N} \left(\hat{\theta}_i - \theta\right)^2}{N}},\ bias(\hat{\theta}) = \hat{\theta} - \theta, e = \frac{\left|bias(\hat{\theta})\right|}{|\theta|},$$

$$f = \frac{RMSE(\hat{\theta})}{|\theta|},$$

where $\hat{\theta}_i$ is the estimator of parameter θ determined in a single simulation and N means the number of repetitions simulation.

3.1. Example 1

This data set gives the gross national product (GNP) per capita in 1996 for various countries as well as their estimated carbon dioxide (CO2) emission per capita for the same year. The data are available in the mixtools package in R. This data frame consists of 28 countries and the following variables:

- GNP – The gross national product per capita in 1996.
- CO2 – The estimated carbon dioxide emission per capita in 1996.

As an example, we fit 2-component model to the GNP data shown in Figure 2. First we use the least squares method for an implicit interdependence applying the function optim in R. The least squares regression lines are given by:

$$\widehat{CO2}_1 = 0.83 \cdot GNP - 4.09 \text{ , } \widehat{CO2}_2 = -0.06 \cdot GNP + 9.51$$

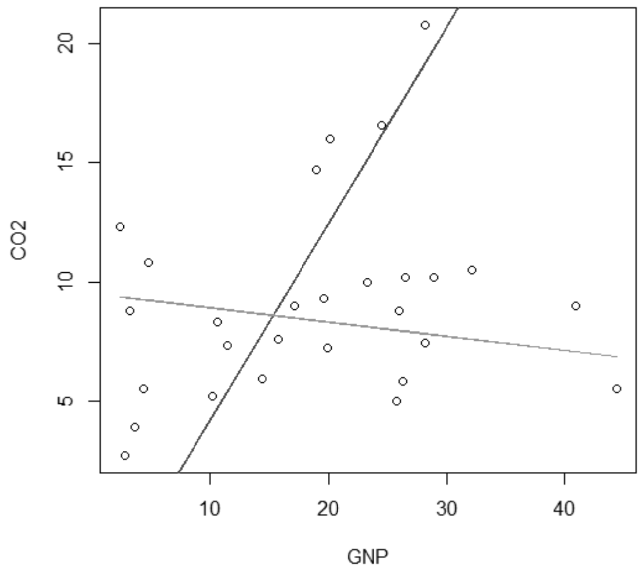

Figure 2. Plot of y versus x with the fitted least squares regression lines

Source: Own calculations.

We estimate the parameters in a mixture regression using maximum likelihood estimation (MLE) method in mixtools R package applying regmixEM function. The function regmixEM will be used for fitting a 2-component mixture of regressions by an EM algorithm:

Table 1. Parameters estimates MLE method

	comp 1	comp 2
$\hat{p}$	0.75	0.25
$\hat{\beta}_1$	8.67	1.41
$\hat{\beta}_2$	-0.02	0.68
$\hat{\sigma}$	2.04	0.81

Source: Own calculation.

The MLE regression lines are given by:

$$\widehat{CO2}_1 = 0.68 \cdot GNP + 1.41, \quad \widehat{CO2}_2 = -0.02 \cdot GNP + 8.67$$

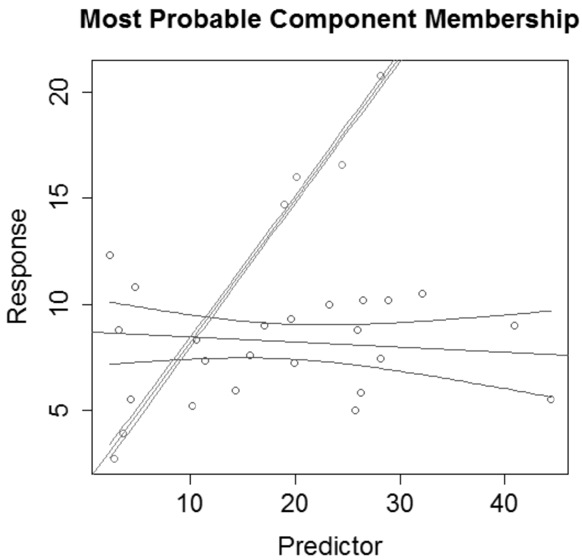

Figure 3. Plot of y versus x with the fitted MLE regression lines

Source: Own calculations.

3.2. Example 2

We randomly generated data (100 points-two straight lines) 10 000 times in R with the following parameters: $a = 2$, $b = 1$, $c = -1$, $d = 30$, $N = 10000$, $n = 100$.

The values of x was generated from the uniform distribution U(0,20):

$$y_1 = 2x + 1 + \varepsilon, y_2 = -x + 30 + \varepsilon, \varepsilon \cong N(0,2)$$

Table 2. Comparison both methods using bias and root mean square error (Example 2)

p = 0.5	Least squares method for an implicit interdependence				Maximum likelihood method			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	0.02	0.01	0.06	0.03	0.01	0.005	0.05	0.025
$\hat{b}$	-0.19	0.19	0.72	0.72	0.01	0.01	0.58	0.58
$\hat{c}$	0.02	0.02	0.06	0.06	-0.01	0.01	0.05	0.05
$\hat{d}$	0.19	0.006	0.73	0.024	-0.02	0.0006	0.59	0.019

Table 2 connt.

p = 0.7	**Least squares method for an implicit interdependence**				**Maximum likelihood method**			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	0.03	0.015	0.06	0.03	0.01	0.005	0.04	0.02
$\hat{b}$	-0.33	0.33	0.77	0.77	0.01	0.01	0.6	0.6
$\hat{c}$	-0.02	0.02	0.08	0.08	-0.01	0.01	0.07	0.07
$\hat{d}$	-0.18	0.006	0.99	0.033	0.03	0.001	0.78	0.026
p = 0.8	**Least squares method for an implicit interdependence**				**Maximum likelihood method**			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	0.04	0.02	0.06	0.03	0.01	0.005	0.04	0.02
$\hat{b}$	-0.38	0.38	0.8	0.8	0.01	0.01	0.62	0.62
$\hat{c}$	-0.07	0.07	0.14	0.14	0.01	0.01	0.09	0.09
$\hat{d}$	-0.75	0.025	1.72	0.057	0.03	0.001	1.01	0.034

Source: Own calculations.

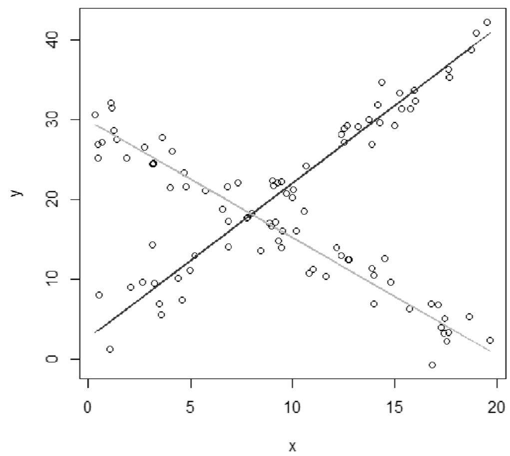

Figure 4. Plot of y versus x with the fitted least squares regression lines; p = 0.5

Source: Own calculations.

The least squares regression lines are given by:

$$\hat{y}_1 = 2.02 \cdot x + 0.81, \quad \hat{y}_2 = -1.02 \cdot x + 30.19$$

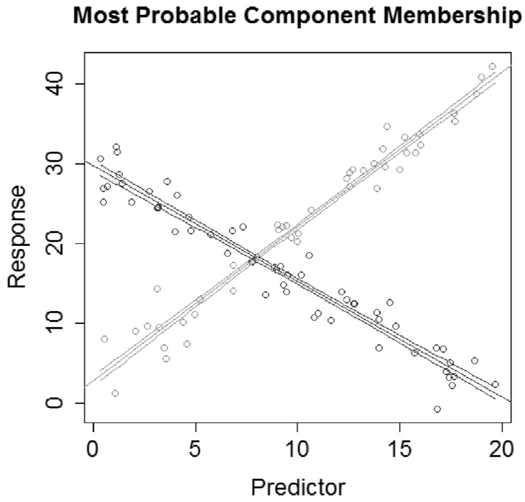

Figure 5. Plot of y versus x with the fitted MLE regression lines; p = 0.5

Source: Own calculations.

The MLE regression lines are given by:

$$\hat{y}_1 = 2.01 \cdot x + 1.01, \quad \hat{y}_2 = -0.99 \cdot x + 29.98$$

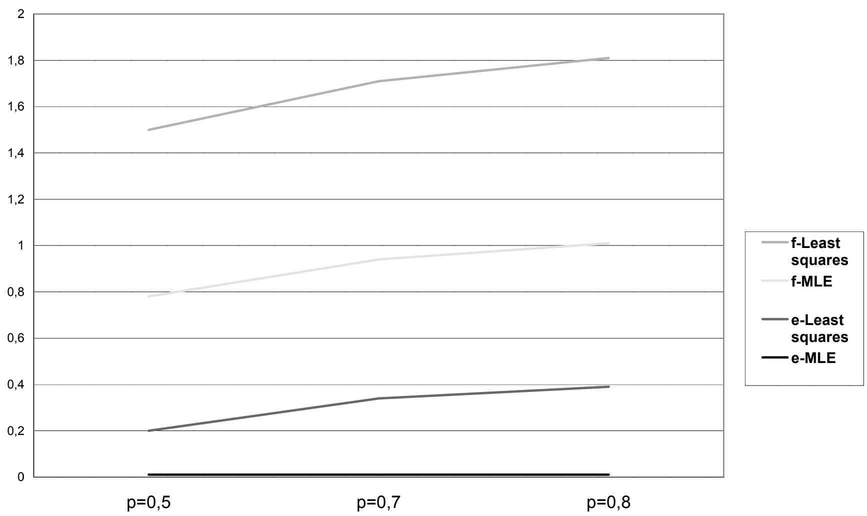

Figure 6. Comparing the accuracy of estimation depending on p for the parameter b (Example 2)

Source: Own calculations.

3.3. Example 3

We randomly generated data (100 points-two straight lines) 10 000 times in R with the following parameters: a = 1.5, b = 1, c = 0.5, d = 0.5, N = 10000, n = 100. The values of x was generated from the uniform distribution U(0,20):

$$y_1 = 1.5x + 1 + \varepsilon, y_2 = 0.5x + 0.5 + \varepsilon, \varepsilon \cong N(0,2)$$

Table 3. Comparison both methods using bias and root mean square error (Example 2)

p = 0.5	**Least squares method for an implicit interdependence**				**Maximum likelihood method**			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	0.03	0.02	0.08	0.053	0.01	0.007	0.05	0.033
$\hat{b}$	0.68	0.68	1.23	1.23	0.03	0.03	0.58	0.58
$\hat{c}$	0.02	0.04	0.09	0.18	-0.01	0.02	0.06	0.12
$\hat{d}$	-0.25	0.5	1.24	2.48	0.05	0.1	0.76	1.52
p = 0.7	**Least squares method for an implicit interdependence**				**Maximum likelihood method**			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	-0.04	0.026	0.08	0.053	0.01	0.007	0.05	0.033
$\hat{b}$	0.97	0.97	1.34	1.34	0.04	0.04	0.6	0.6
$\hat{c}$	0.02	0.04	0.1	0.2	-0.01	0.02	0.08	0.16
$\hat{d}$	-0.17	0.34	1.33	2.66	0.04	0.08	1.11	2.22
p = 0.8	**Least squares method for an implicit interdependence**				**Maximum likelihood method**			
	bias	e	RMSE	f	bias	e	RMSE	f
$\hat{a}$	-0.04	0.026	0.08	0.053	0.01	0.005	0.05	0.033
$\hat{b}$	1.06	1.06	1.42	1.42	0.04	0.04	0.6	0.6
$\hat{c}$	0.02	0.04	0.14	0.28	-0.01	0.02	0.11	0.22
$\hat{d}$	0.2	0.4	1.69	3.38	0.08	0.16	1.51	3.02

Source: Own calculations.

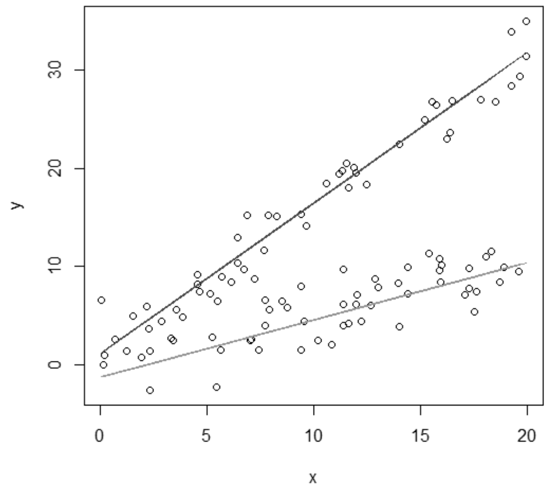

Figure 7. Plot of y versus x with the fitted least squares regression lines; p = 0.5

Source: Own calculations.

The least squares regression lines are given by:

$$\hat{y}_1 = 1.47 \cdot x + 1.68,\ \hat{y}_2 = 0.48 \cdot x + 0.25$$

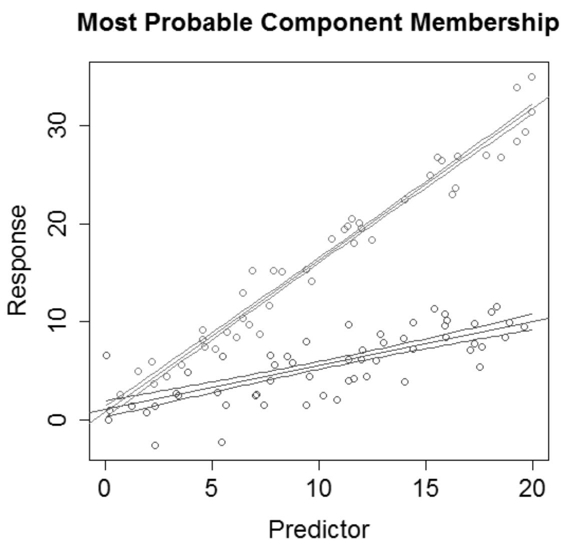

Figure 8. Plot of y versus x with the fitted MLE regression lines; p = 0.5

Source: Own calculations.

The MLE regression lines are given by:

$$\hat{y}_1 = 1.51 \cdot x + 1.03, \quad \hat{y}_2 = 0.49 \cdot x + 0.55$$

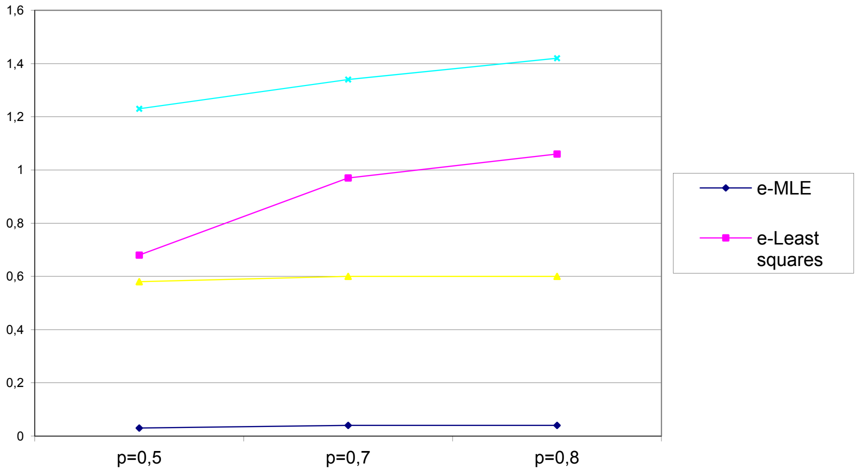

Figure 9. Comparing the accuracy of estimation depending on p for the parameter *b* (Example 2)

Source: Own calculations.

3.4. Bootstrapping for standard errors

With likelihood methods for estimation in mixture models, it is possible to obtain standard error estimates by using the inverse of the observed information matrix when implementing a Newton-type method. However, this may be computationally burdensome. An alternative way to report standard errors in the likelihood setting is by implementing a parametric bootstrap. Efron and Tibshirani [1993] claim that the parametric bootstrap should provide similar standard error estimates to the traditional method involving the information matrix.

In a mixture-of-regressions context, a parametric bootstrap scheme may be outlined as follows:

1. Use function regmixEM to find a local maximizer $\hat{\theta}$ of the likelihood or use function optim to find $\hat{\theta}$ applying the least squares method for an implicit interdependence.
2. For each x_i simulate a response value y_i^* from the mixture density $g_{\hat{\theta}}(\cdot \mid x_i)$.

3. Find a parameter estimate $\tilde{\theta}$ for the bootstrap sample using function regmixEM or the least squares method for an implicit interdependence.
4. Use some type of check to determine whether label-switching appears to have occurred – and if so – correct it.
5. Repeat steps 2 through 4 B times to simulate the bootstrap sampling distribution of $\hat{\theta}$.
6. Use the sample covariance matrix of the bootstrap sample as an approximation to the covariance matrix of $\hat{\theta}$.

Table 4. Bootstrap standard error estimates for parameters in Example 2; B = 10 000

	Least squares method for an implicit interdependence			**Maximum likelihood method**		
	original	relative error	standard error	original	relative error	standard error
$\hat{a}$	2	0.035	0.07	2	0.03	0.06
$\hat{b}$	1	0.6	0.6	1	0.71	0.71
$\hat{c}$	-1.0	0.09	0.09	-1.0	0.049	0.049
$\hat{d}$	30	0.045	1.35	30	0.018	0.54

Source: Own calculations.

In this example, $\hat{\theta}_i^*$ are bootstrap copies of $\hat{\theta}$:

$$relative\,error = \frac{\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \bar{\hat{\theta}}^*\right)^2}}{|\theta|},\ \bar{\hat{\theta}}^* = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^*\ .$$

Table 5. Bootstrap standard error estimates for parameters in Example 3; B = 10 000

	Least squares method for an implicit interdependence			**Maximum likelihood method**		
	original	relative error	standard error	original	relative error	standard error
$\hat{a}$	1.5	0.047	0.07	1.5	0.04	0.06
$\hat{b}$	1.0	0.83	0.83	1.0	0.62	0.62
$\hat{c}$	0.5	0,2	0.1	0.5	0.16	0.08
$\hat{d}$	0.5	2.48	1.24	0.5	1.96	0.98

Source: Own calculations.

Conclusions

In this article, we compare accuracy of two methods of estimation for the parameters of a mixture of linear regressions in the case when mixture components are normally distributed. The relative bias and relative root mean square error of estimators obtained with the method of maximum likelihood they are smaller than when using the least squares method for an implicit interdependence. The relative bias and relative root mean square error of estimators obtained by both method is the smallest when $p = 0.5$.

Bootstrapping standard errors and relative errors of estimation were smaller for the method of maximum likelihood. The advantage of the least squares method for an implicit interdependence is shorter computation time. The least squares method for an implicit interdependence does not require determine the distribution of the component mixture.

References

Antoniewicz R. (1988), *Metoda najmniejszych kwadratów dla zależności niejawnych i jej zastosowanie w ekonomii*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.

Benaglia T., Chauveau D., Hunter D.R., Young D. (2009), *Mixtools: an R Package for Analyzing Finite Mixture Models*, "Journal of Statistical Software", Vol. 32, Issue 6.

McLachlan G.J., Peel D. (2000), *Finite Mixture Models*, John Wiley & Sons, New York.

Dogru F.Z., Arslan O. (2014), *Robust Mixture Regression Based on the Skew t Distribution*, Ankara University, Ankara.

De Veaux R.D. (1989), *Mixtures of Linear Regressions*, "Computational Statistics and Data Analysis", No. 8, s. 227-245.

Dempster A.P., Laird N.M., Rubin D.B. (1977), *Maximum Likelihood from Incomplete Data Via the EM Algorithm*, "Journal of the Royal Statistical Society B", No. 39(1), s. 1-38.

Efron B., Tibshirani R. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, London.

Lindsay B.G. (1995), *Mixture Models: Theory, Geometry and Applications*, Institute of Mathematical Statistics, Hayward.

Pruska K. (1992), *Metoda największej wiarygodności a regresja przełącznikowa*, „Folia Oeconomica", nr 117, s. 107-130.

Quandt R.E. (1972), *A New Approach to Estimating Switching Regressions*, "Journal of the American Statistical Association", No. 67(338), s. 306-310.

Quandt R.E., Ramsey J.B. (1978), *Estimating Mixtures of Normal Distributions and Switching Regressions*, "Journal of the American Statistical Association", No. 73(364), s. 730-738.

ESTYMACJA PARAMETRÓW REGRESJI MIESZANKI DWUWYMIAROWYCH ROZKŁADÓW PRAWDOPODOBIEŃSTWA

Streszczenie: Do estymacji parametrów mieszanek regresji stosujemy dwie metody: metodę największej wiarygodności oraz metodę najmniejszych kwadratów dla zależności niejawnych. Najbardziej popularną metodą polegającą na maksymalizacji funkcji wiarygodności jest algorytm EM. Metoda najmniejszych kwadratów dla zależności niejawnych polega na rozwiązaniu układu równań nieliniowych. Najczęściej stosowaną metodą estymacji parametrów mieszanek regresji jest metoda największej wiarygodności. W artykule pokazano możliwość zastosowania innej metody najmniejszych kwadratów dla zależności niejawnych. Obie metody porównujemy symulacyjnie, używając obciążenia estymatora, pierwiastka błędu średniokwadratowego estymatora oraz bootstrapowe błędy standardowe.

Słowa kluczowe: mieszanki regresji, algorytm EM, metoda najmniejszych kwadratów dla zależności niejawnych.