Proceedings of the 11th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'11), 19–22 October 2011, Riga, Latvia, p. 28-35. ISBN 978-9984-818-46-7
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

USE OF R ENVIRONMENT FOR A FINDING PARAMETERS NONLINEAR REGRESSION MODELS

Victor Lyumkis, Andris Plotka, Aleksandr Krivchenkov

Transport and Telecommunication Institute
Lomonosova str. 1, Riga, Latvia
E-mail: vlyumkis@yahoo.com, andris.plotka@gmail.com, aak@tsi.lv

In the present work the actual problem of finding of parameters nonlinear regression models on sample data is considered. In work the accent on the practical approach of the solving of similar problems is done, using powerful means of modern package R. It is possible to note, that the method of the least squares remains the major tool for an estimation of parameters and resulted codes allow realizing rather simply this method in R environment. On the other hand, the opportunity of application of kernel regression for data smoothing is discussed, the optimality of such approach shows at greater sample sizes. In the article a lot of practical examples with concrete methodology of their realization in package R is considered

Keywords: R environment**,** nonlinear regression analysis, kernel estimations

1. Introduction

Still comparatively recently the problem of finding the estimations of the parameters of nonlinear regression models according to the concrete observations remained completely labour-consuming. At present with the advent of contemporary package means solution of similar problems to a considerable extent simplified. In this work the discussion will deal with the use of the classical method of least squares ([MLS]) on environment of the packet R [6], whose powerful computational and graphic merits are completely well-known. We will discuss a number of the practical examples to the realization of nonlinear regression models, attempting to follow some methodological recommendations, presented in the book [3].

From other side, the finding of the smoothing out regression in a number of cases can be realized with the aid of the so-called kernel regression [4]. This approach also is discussed in the work. In this connection let us note in more detail that in opposed to the classical method of the least squares, considered above, it is possible to paraphrase a little in some cases required regression equation of relationship between dependent variable Y and a vector of explaining variables X and to reduce to model of a kind

$$Y = m(X) + \varepsilon, \tag{1}$$

where ε – some random variable having at everyone X=x distribution with a mathematical expectation $M(\varepsilon) = 0$ and a variation $D(\varepsilon) = \sigma^2$, m(X) – some smooth function.

By definition the conditional average of a continuous random variable can be presented thus [4]

$$M(Y|X = x) = \int yg(y/x)dy = \int y\frac{f(x,y)}{f(x)}dy = m(x), \tag{2}$$

where g (y/x) – conditional density of random variable Y at fixed value X=x, and f (x, y) joint density of random variables X, Y.

The estimation of a conditional average turns out replacement of unknown functions of density joint and marginal distributions, f (x, y) and f (x), their "good" estimations. Thus "good" estimation of a conditional average is the kernel estimation (3) which detailed design is described, for example, in works [4, 5]. It allows reducing an estimation of conditional average Y at everyone x to an estimation of a kind:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K(\frac{X_i - x}{h})}{\sum_{i=1}^{n} K(\frac{X_i - x}{h})}, \tag{3}$$

where K (z) – value, for example, function of density of standard normal distribution in a point z, X_i,Y_i ,i=1,…,n, known sample of relationship between Y and X.

The concrete examples of the application of kernel regression are also discussed in this work.

2. Basic Methodological Recommendations of the Use of Classical MLS in Environment R. Example 1.

As it is known, the most important question in many fields of science is the detection of the connection between two variable X and Y (the X and Y, generally speaking, some continuous random variables). Frequently succeeds in postulating the so-called regression equation of relation between dependent variable Y and independent variable of the X form

$$Y = f(X,\theta) + \varepsilon, \tag{4}$$

where ε – a certain random variable, which has with each X=x distribution with the mathematical expectation $M(\varepsilon) = 0$ and variation $D(\varepsilon) = \sigma^2$; $f(X,\theta)$ – a certain, in the general case known, fairly complicated nonlinear, smooth function, which depends on the vector of the explaining variables X=$(x_1, x_2, ...x_k)$ and the vector of the parameters $\theta = (\theta_1, \theta_2, ..., \theta_p)$.

Standard method of the finding of the vector of the unknown parameters as it is known [1], it is reduced to the minimization of the sum of the squares of discrepancies – to use of the method of least squares ([MLS]), which is reduced to the minimization of the function of the form

$$RSS(\theta) = \sum_{i=1}^{n} (Y_i - f(x_i,\theta)^2, \tag{5}$$

where n – number of observations of dependent variable Y.

To find estimation $\hat{\theta}$ on MLS we should differentiate (5) on θ and equate the left parts to 0 for reception of the so-called system p normal equations. Difficulties of the solving of received system of the normal equations are well-known, and for its solving it is necessary to use iterative methods, moreover difficulties are aggravated with that there can be a set of stationary values of function S(θ).

In this connection the major meaning gets a correct choice of initial values by search of a global minimum of system (5). And in it the essence of the first methodical reception consists – a correct finding initial approach convenient graphic means in package R. We shall consider in detail an example 1, a mention about which is available in the book [1]. In Table 1 the given dependences of variable Y from x are presented in the assumption that true dependence looks like

$$Y = \alpha + (0.49 - \alpha)\exp(-\beta(x - 8)) + \varepsilon. \tag{6}$$

Table 1. Dependence Y from x for model (6)

x	10	11	12	13	14	15	16	17	18	19	20	22	24
Y	**0.48**	0.47	0.46	0.45	0.48	0.44	0.43	0.44	0.43	0.45	0.42	0.414	0.413

Continuation. **Table 1**

x	26	28	30	32	34	36	38	40
Y	0.48	0.42	0.40	0.401	0.398	0.397	0.396	0.39

It is necessary by means of method of the least squares to receive an estimation of parameters of this nonlinear model, thus it is possible to emphasize, that search of parameters of model is carried out on enough small sample that happens is characteristic for of some practical problems. Analysing function, it is visible, that at x=40, exp (-32*β), then $\alpha \approx 0.39$. For a finding of starting value of parameter, we shall substitute in the set function of value: α=0.39, x=20, y=0.42 also we shall receive prospective initial value β =0.1. The similar way of a choice of initial parameters is rather important. However, for similar problems in package R the convenient graphic interface connected with visualization of smoothing of data is offered. Realization of the specified code in the environment of package R is presented on Figure 1.

```
Observ1<-data.frame(x<-
c(10,11,12,13,14,15,16,17,18,19,20,22,24,26,28,30,32,34,36,38,40),
y<-c(0.48,0.47,0.46,0.45,0.48,0.44,0.43,
0.44,0.43,0.415,0.42,0.414,0.413,0.41,0.405,0.40,0.401,0.398,0.397,0.3
96,0.39))
Model1 <- function (x,alpha,beta)
{
      alpha+(0.49-alpha)*exp(-beta*(x-8))
}

plot(y~x,lwd=5)
curve(Model1(x, alpha = 0.39, beta = 0.1),col="red", add = TRUE,lty=2)
curve(Model1(x, alpha = 0.4, beta = 0.15),col="green", add = TRUE,lty=3)
curve(Model1(x, alpha = 0.41, beta = 0.1),col="blue", add = TRUE,lty=4)
```

Figure 1. Search of parameters for a choice of the best smoothing sample data for an example 1 in environment R

The correlation field of initial data and results of graphic smoothing are presented on Figure 2. From figure it is visible, that the best initial parameters for smoothing are provided with a choice of a red line (average of three), where α =0.39,β =0.1

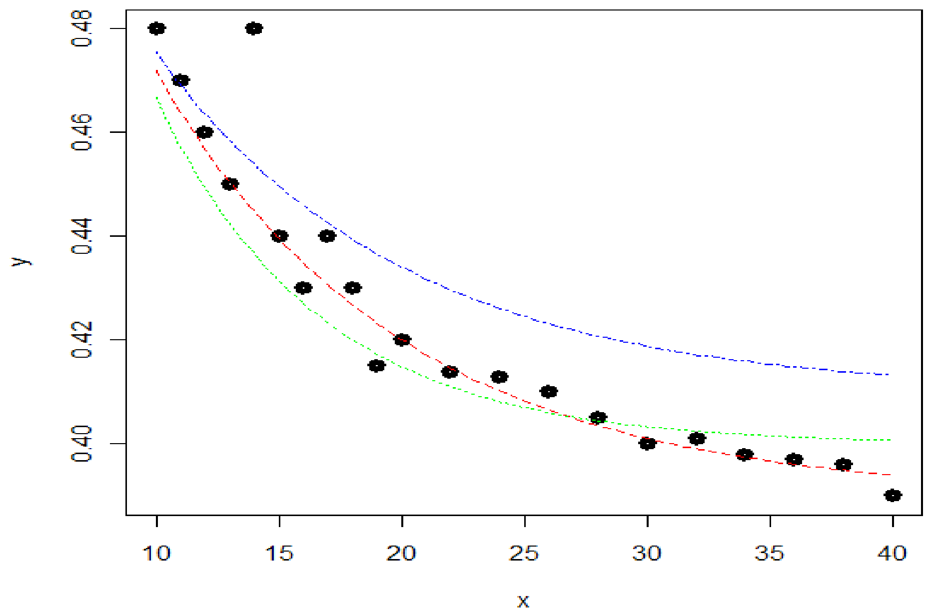

Figure 2. Correlation field and results of smoothing for example 1

Let's notice, in particular, that the specified search of parameters with graphic display of results is the first convenient methodical recommendation at selection of parameters for nonlinear regress.

The second methodical aspect closely connected with the previous one, is "greed approach" when on the set grid of parameters gets out those their values which give the minimal value to residual sum RSS (θ) (5). For reception of more exact values we spend search by a grid, changing values α from 0.3 up to 0.5 with step 0.01, and values β – from 0 up to 0.5, too with step 0.01.

Search on a grid in R environment is done as follows:

```
grid.Observ1 <- expand.grid(list (alpha = seq(0.3,
      0.5, by = 0.01), beta = seq(0, 0.5, by = 0.01)))

Observ1.m2a <- nls2(y ~ Model1(x,
alpha, beta), data = Observ1,
start = grid.Observ1, algorithm = "brute-force")
```

Results of the done analysis allow to choose optimum initial values α =0.39; β =0.09, thus the counted up sum RSS (α, β) =0.0014.

The final choice of parameters after is found initial approach, is carried out within the limits of known iterative standard function of a package nls () (it is the third recommendation) and results in the following:

```
Observ1.m1 <- nls(y ~ Model1(x,alpha,beta),
      data=Observ1, start = list(alpha=0.39,
      beta=0.09), trace=TRUE)
 α=0.383; β=0.08 ; RSS(α, β)=0.0135;
```

Summing up to discussion to value of discrepancy RSS, which can be useful at comparison, for example, various models of smoothing, special function deviance () is developed. In our case its application is presented below and gives result which was already discussed:

deviance (Observ1.m1)
Received value of discrepancy RSS is equal to 0.00135.

3. Discussion of an Example 2

Let similarly to the previous example is considered the nonlinear model of a kind

$$Y = \theta_1 * (1 - \exp(-x * \theta_2)) + \varepsilon \, . \tag{7}$$

On the sample data presented in Table 2, it is necessary to estimate parameters of model using method of the least squares

Table 2. Dependence Y from x for model (7)

x	1	2	3	4	5	7	8	9	10	11
Y	**9**	9	16	20	21	22	23	24	24	25

Using some search with graphic visualization of results, we shall find in R environment reasonable initial values of required parameters. It is provided with a following code in the environment of package R:

```
observ2<-data.frame(x<-c(1,2,3,4,5,7,8,9,10,11),
y<-c(9,9,16,20,21,22,23,24,24,25))

Model2 <- function (x,teta1,teta2)
{
      teta1*(1-exp(-teta2*x))
}
plot(y ~ x,lwd=5)
curve(Model2(x, teta1 = 25, teta2 = 0.3),
      col="red",lty=1, add = TRUE)
curve(Model2(x, teta1 = 25.26, teta2 = 0.327),
      col="brown",lty=2, add = TRUE)
curve(Model2(x, teta1 = 26, teta2 = 0.25),
      col="blue",lty=3, add = TRUE)
curve(Model2(x, teta1= 24, teta2 = 0.3),
      col="green",lty=4, add = TRUE)
```

The correlation field of initial data and results of graphic smoothing are presented on Figure 3. From figure it is visible, that the best initial parameters for smoothing are provided with a choice first of four where θ1=25.26, θ2=0.327, the choice of these parameters is explained below.

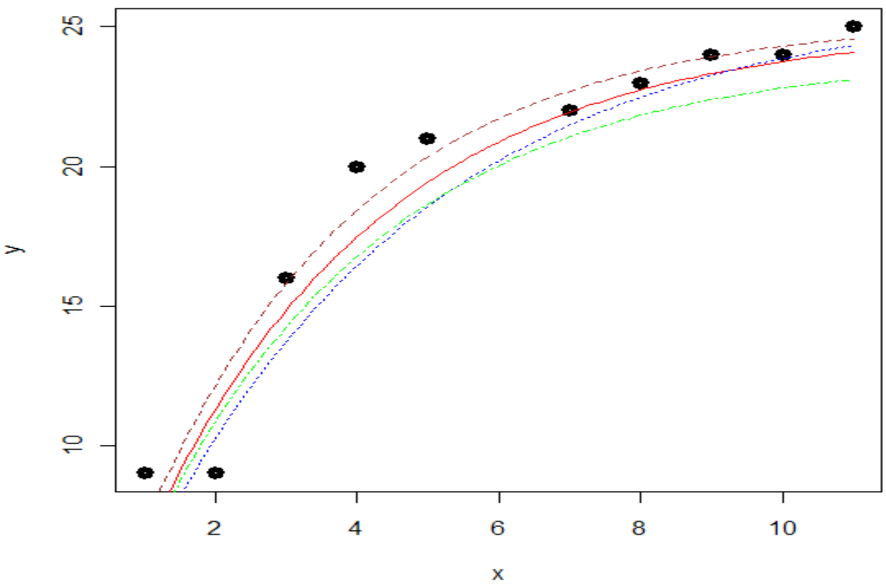

Figure 3. Correlation field and results of smoothing for example 2

As well as in the previous example, for reception of more exact values we spend search by a grid, changing values θ1 from 20 up to 25 with step 0.01, and values θ2 – from 0 up to 0.5, too with step 0.01:

```
grid.Observ1 <- expand.grid(list (teta1 = seq(20,
      30, by = 0.01), teta2 = seq(0, 0.5, by = 0.01)))
```

```
Results of the done analysis allow to choose optimum initial values of
parameters θ1=25.9;θ2=0.3, thus the counted up sum RSS (θ1, θ2)
=18.46.
The final choice of parameters after are found initial approach, is
carried out within the limits of known iterative standard function of
a package nls () and results in the following
Observ1.m1 <- nls(y ~ Model2(x,teta1,teta2),
      data=observ2, start = list(teta1=25.9,
      teta2=0.3), trace=TRUE)
θ1=25.26; θ2=0.327; RSS(θ1,θ2)=17.4
```

4. Comments to Alignment of Models with Several Predictive Variables

Presence of two or more predictive variables does not change the basic idea of a finding of an optimum estimation θ in expression (5). It is clear, that graphic interpretation of a choice of an optimum smoothing curve becomes complicated, however software for minimization the discrepancy RSS(θ1, θ2) remain former. We shall consider an example, the mention about which is available in work [1]. Let on known sample data x1, x2, Y it is necessary to estimate a method of the least squares parameters θ1, θ2, θ3 for the model

$$Y = \theta_1 * \theta_3 * x_1 / (1 + \theta_1 * x_1 + \theta_2 * x_2) + \varepsilon . \tag{8}$$

All the detailed results connected with a finding of required parameters, are presented below in program listing on Figure 4.

```
Observ4 <- data.frame(x1<-c(1,2,1,2,0.1,3,0.2,3,0.3,3,3,0.2,3),
x2<-c(1,1,2,2,0,0,0,0,0,0.8,0,0,0.8),
y<-
c(0.126,0.219,0.076,0.126,0.186,0.606,0.268,0.614,0.318,0.298,0.509,
0.247,
0.319))
Model4 <- function (x1,x2,t1,t2,t3)
{
      t1*t3*x1/(1+t1*x1+t2*x2)
}
#plot(y ~ x)
grid.Observ4 <- expand.grid(list (t1 = seq(2.9,
      4, by = 0.01), t2 = seq(12.2, 12.8, by = 0.01), t3 = seq(0.62,
0.7, by = 0.05)))
Observ1.m2a <- nls2(y ~ Model4(x1,x2,
t1,t2,t3), data = Observ4,
start = grid.Observ4,algorithm = "brute-force")
Observ1.m2 <- nls(y ~ Model4(x1,x2,t1,t2,t3),
      data=Observ4, start = list(t1=3.57,t2=12.7,t3=0.6), trace=TRUE)
Observ1.m2
Nonlinear regression model
  model:  y ~ Model4(x1, x2, t1, t2, t3)
   data:  Observ4
     t1      t2      t3
 3.5691 12.7959  0.6295
 residual sum-of-squares: 0.00788
```

Figure 4. Program code of a finding of parameters for model 8

From Figure 4 it is visible, that for the entered data required parameters can be found as it was done in the previous cases. In the beginning are found reasonable approach of parameters of model use of function nls2 (). Received approach with use of iterative function R nls () allow to find optimum values of parameters θ1=3.57; θ2=12.79; θ3=0.63, discrepancy RSS(θ1,θ2,θ3) is equal 0.0078. We shall notice that scanning of this value of discrepancy in wide enough range only has confirmed the correctness found estimations of parameters.

5. Use of Kernel Approach for Building a Nonlinear Regression Model

We have already noted in the introduction of the possibility and desirability of obtaining regression smoothing using a kernel approach. Let us consider in detail the implementation of this approach in an environment R for a particular model

$$Y = a - b * \exp(-c * x) + \varepsilon \ . \quad (9)$$

Based on this model, simulate a sample of n = 100 with parameters a = 115, b = 118, c = 0.15 for the average values of Y and add a "noisy" [4]. As is well known for the software environment R was designed package np, which contains an easy to use and open source code for kernel estimation. Package np, is composed of module npregbw, which is used for selecting the optimal window width in kernel regression [4] and the module npreg, and used for most kernel regression. Note that the module npregbw can select the different cores smoothing – Epanechnikov, Gaussian kernel and the uniform. Figure 5 shows the implementation of the choice of kernels.

```
bw.gaussian <- npregbw(y~x,ckertype="gaussian",tol=.1,ftol=.1)
bw.epanechnikov <- npregbw(y~x,ckertype="epanechnikov",tol=.1,ftol=.1)
bw.uniform <- npregbw(y~x,ckertype="uniform",tol=.1,ftol=.1)
```

Figure 5. Choosing the smoothing kernel in the package npregbw

Using the referred above code in the work was performed the nonparametric smoothing using different kernels. Figure 6 shows the correlation field generated data, the true DGP (data generating process) and the results of graphical smoothing.

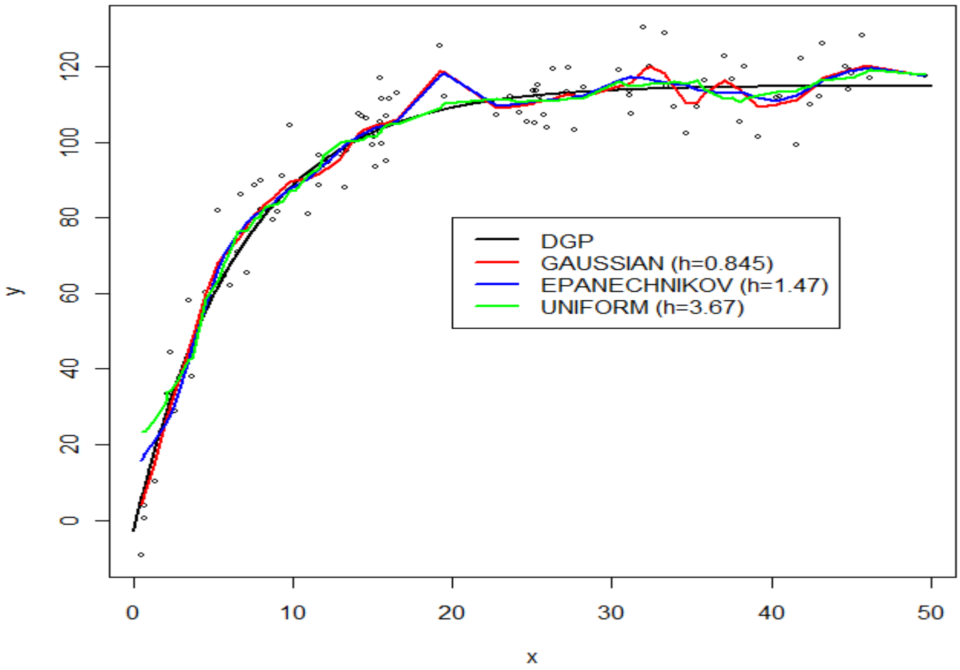

Figure 6. Correlation field and result in the application of smoothing kernels Epanechnikov, Gaussian and uniform, n = 100

As Figure 6 shows, the choice of different kernels affects the final result of smoothing of the sample. We also see that the kernel smoothing rather good approximation of our data, even when the sample, n = 100. With the increase in the number of observations or in the case when the parametric method is difficult to implement, kernel smoothing methods can be regarded as a great alternative for implementing high-quality regression analysis. We demonstrate this by gradually increasing the sample formed using the model (9). To do this will generate an additional sample of size n = 200, n = 500, n = 2000. When using a Gaussian kernel smoothing kernel and use a specific type of smoothing parameter – a locally constant. Figure 7 compares the results of the parametric kernel smoothing and smoothing using a least squares method.

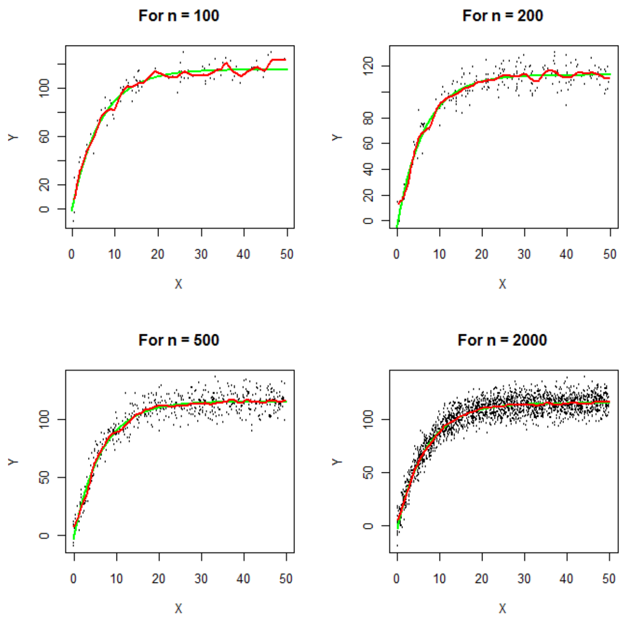

Figure 7. Comparison of kernel and parametric regression smoothing methods, n = 100, 200, 500, 2000

Commenting on the previous figure, we see that the asymptotic sample difference between the kernel approach and an approach based on the method of least squares is leveled. Similarly, we can see that with increasing size n of sampling, all kernel regression estimates, regardless of the form of the kernel is also given almost equivalent result of smoothing. With a sample n = 2000 is carried out smoothing the data using different kernels, and on Figure 8 shows the compliant result of smoothing.

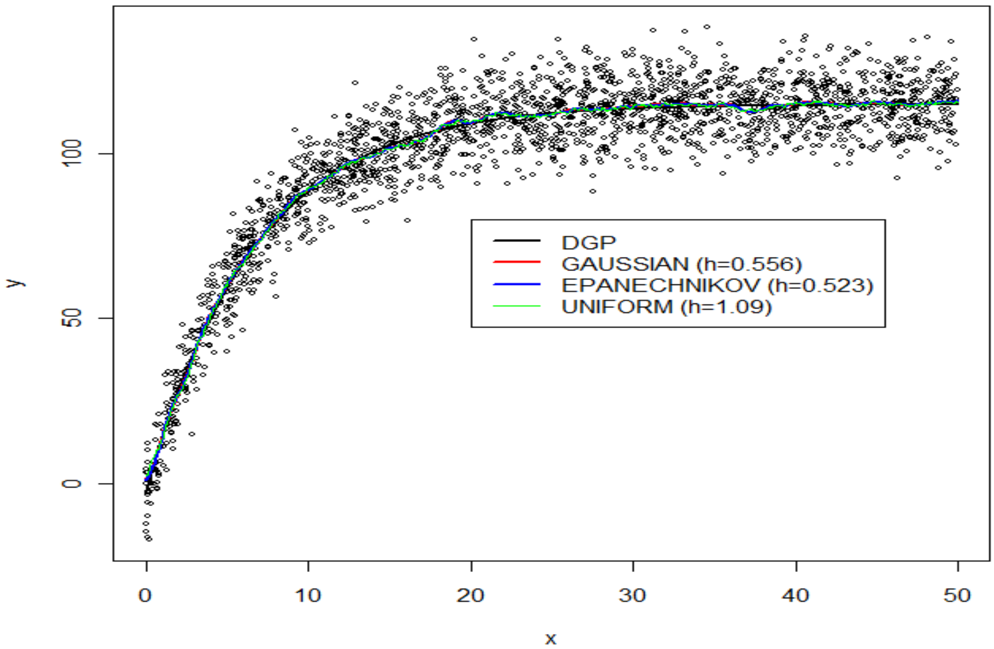

Figure 8. Correlation field and result in the application of smoothing kernels Epanechnikov, Gaussian and uniform, n = 2000

6. Conclusions

For practical realization of the parametric nonlinear least squares regression model remains the central method of finding the parameter estimates. However, among the ready-made software using least squares for finding the parameters of the models can only recommend freeware package R. This paper shows a simple method of finding the parameters of nonlinear regression by means of the package R for some specific examples; also have code in an environment R for their achievement. Without exaggeration, the advantages can be assumed that any one-dimensional non-linear regression model can be easily implemented by the proposed method in the package R.

In the case of the need to build non-parametric nonlinear regression investigated the possibility of using kernel regression. A custom implementation of the kernel regression using specific examples in the medium R is showed. It has been observed that the asymptotic sample size is constructed models, regardless of the form of the kernel and the choice of other parameters are practically identical. Moreover, they practically coincide with the regression obtained by least squares.

References

1. Draper, N., Smith, H. *Applied Regression Analysis. Third ed.* New York: John Wiley and Sons, 1998. 706 p.
2. Bates, D.M., Watts, D.G. *Nonlinear regression analysis and its applications*. New York: John Wiley and Sons, 1988. 365 p.
3. Ritz,J. C., Streibig, C. *Nonlinear regressions with R*. New York: Springer, 2008. 144 p.
4. Racine, Jeff & Li. Nonparametric estimation of regression functions with both categorical andcontinuous data, *Journal of Econometrics,* Vol. 119(1), March 2004, pp 99-130.
5. Härdle, W., Muller, M., Sperlich, S, Werwatz, A. *Nonparametric and Semiparametric Models.* Berlin – Heidelberg-New York: Springer, 2004. 328 p.
6. *R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing.* Vienna, Austria, 2005. ISBN 3-900051-07-0, URL: http://www.R-project.org