20th International Scientific Conference AMSE
Applications of Mathematics and Statistics in Economics 2017
Szklarska Poręba, Poland 30 August 2017 – 3 September 2017

MODELLING THE FORCE OF MORTALITY USING LOCAL POLYNOMIAL METHOD IN R

SAMUEL HUDEC
Matej Bel University, Faculty of Natural Sciences, Department of Mathematics,
Tajovského 40, Banská Bystrica, Slovak Republic
email: samuel.hudec@umb.sk

Abstract

This paper discusses the use of non-parametric regression approach as a new view of modelling the force of mortality which is an important and fundamental concept in modelling the future lifetime. The local polynomial method uses robustness ideas from the linear regression combined with the local fitting ideas from the kernel methods. Our topic is to compare this approach with Gompertz Makeham's law based method. Differences are presented throught the amounts of selected product of supplementary pension saving constitutes the third pillar of the pension system in Slovakia. All of the computations for this paper was undertaken in statistical software R.

Key words: *force of mortality, local polynomial method, Makeham's law, non-parametric regression*

JEL Codes: *C14, C51, G22, I13, H51*

DOI: 10.15611/amse.2017.20.17

1. Introduction

One of the important tasks in actuarial science is the force of mortality that represents instantaneous rate of mortality at a certain age measured on an annualized basis. True force of mortality curve (function) is unknown, we are only able to model this true function from collected data represented by the ratio of number of dying to number of living at a current age. Therefore, we get the estimatimation of Survival function based on noise dataset.

In the paper we visualize how to find the true function of the force of mortality. In the last decades of the 20th century, statisticans focused on the parametric model founded by Benjamin Gompertz see (Gompertz 1825), and later updated by William Makeham see (Makeham 1860). Gompertz started with a simple parametric equation which involves basic intuition about the true function also called the Standard Survival model. His exponentially increasing function used to date drew our attention to a difficult restricted optimalization problem. Every statistical software implement Standard Survival model with different solutions and optimalization forms influenced by the start estimations. Also, the choice of a numerical method is important and numerically sensitive. For all the reasons above we made a decision to use the estimations of Standard Survival model follows Špirková et al. (2016).

In this paper we look into the Survival model from non-parametric statistical point of view. Notice that we are in a one dimensional space represented by one regressor (age) and response (ratio of dying to living at a current age) in otherwords, we are in a field of simple regression. Modern non-parametric regression techniques give us a lot of opportunities to find an approximation of the true function, namely kernel smoothers, B-splines, smooth splines, isotonic regression, wavelets etc., see (Faraway 2016). Kernel smoothers achieve flexibility in estimating the true function over the domain (increasing ages) by fitting different but simple model separately at each point of interest or a target point. This is done by using only observations close to the target point (target age) to fit the simple model. Therefore, we could imagine a sliding window from left (lower age) to right (higher age) on domain wherein we compute our simple model (for example average). This window covering area around the targed point is represented by a weighting function or a kernel. The kernels are typically indexed by a parameter that dictates the width of the window. For modelling we use Local polynomial regression which is a combination of kernel methods and linear regression. In other words, in every tagret point we get weighted polynomial function estimation which is numerically stable. Advantages of modelling such way are robustness and correction of bias on regions of curvature of the true function. That involves a combination of average, which is presereved, and soft curve on domain. Also Local polynomial models are used for modelling time series see (Ledolter, 2008).

This paper is divided into four sections and conslusion. In Section 2, we establish the basic essential theory and technical details for modelling the force of mortality via Standard Survival model and Survival model using Local polynomial regression approach. Section 3 includes an application of these methods on life tables published on the website of the Statistical Office of the Slovak Republic and comparison of these methods. In Section 4 there are analysed the amounts of selected product of supplementary pension saving constitutes the third pillar of the pension system in Slovakia with respect to mentioned models of the force of mortality in 2014 whith 1.9% old technical interest rate and 0.7% current technical interest rate (National Bank of Slovakia, 2013, 2015). Moreover, in this part we present confidence intervals as an additional benefit from non-parametric regression as well as we analyse the amounts from "confidence bounds" for the mentioned selected product.

Our aim is to present and compare this new numerically stable method which does not require hard optimalization skills. Also the Local polynomial regression involve a strong intuition about unknown true function as well as the Standard Survival model.

All of the computations for this paper was undertaken in statistical software R. For more details see R website which can be found at (R Core Team 2013). Ploted figures were supported by R library `ggplot2` developed by (Wickham 2016).

2. Basic Concept of Survival Modelling

Let x-th denote a life aged x, where $x > 0$. The death of x-th can occur at any age greater than x, and we can model future lifetime of x-th by continous random variable T_x. We denote true function of the force of moratlity at age x by μ_x using definition as in Dickson et al. (2009).

$$\mu_x = \lim_{dx \to 0^+} \frac{1}{dx} Pr[T_x \leq dx], \tag{1}$$

which can be estimated by ratio $m_x = D_x/P_x$, where D_x is the number of dying at age x and P_x is the number of living at age x. Standard Survival model based on Gompertz law models the estimation of the force of mortality as follows:

$$\hat{\mu}_x = \hat{A} + \hat{B}\hat{c}^x, \tag{2}$$

where $\hat{A}$, $\hat{B}$ and $\hat{c}$ are estimations of parameters and x is a target age of individual. Throughout this paper we illustrate our results on the non-parametric Survival model using Local polynomial regression approach, which models for each target point (target age) x the force of mortality as follows:

$$\hat{\mu}_x = \hat{\alpha}(x) + \sum_{j=1}^{d} \hat{\beta}_j(x)x^j. \tag{3}$$

where $\alpha(x)$ and $\beta_j(x)$ are regression coefficients for each target age x. Equation (3) can by solved as extended least squares problem using formula as in Hastie at al. (2009):

$$\min_{\alpha(x),\beta_j(x),j=1,2,\ldots,d} \sum_{i=1}^{N} K_\lambda(x,x_i)\left[y_i - \alpha(x) - \sum_{j=1}^{d} \beta_j(x)x_i^j\right]^2 \tag{4}$$

where:

- N - sample size, in our case $N = \omega$ maximum age to which a person can live,
- $K_\lambda(x,x_i)$ - kernel or weighting function, which assigns a weight to x_i based on its distanece from x,
- λ - dictates the width of the neighborhood,
- d - polynomial degree, for modelling the force of mortality $d = 2$ is sufficient.

In other words Local polynomial regression solves separate weighted least squares problem at each target point x. Equations (2) and (3) decribe estimations of function μ_x thus if we know some approximation of $\hat{\mu}_x$ from m_x, then we can calculate all the survival probabilities using actuarial notation following Dickson et al. (2009)

$${}_tp_x = \exp\left\{-\int_x^{x+t} \mu_r \,\mathrm{d}r\right\} \approx \exp\left\{-\sum_{r=x}^{x+t} \mu_r\right\}, \tag{5}$$

where ${}_tp_x$ is the probability that individual x-th survives to at least age $x+t$. Particular discrete solutions for each target point (age) x are sufficient in this paper.

3. Data and Results of Analysis

On modelling of the estimation of the true function of the force of mortality, we use life tables which are published on the website of the Statistical Office of the Slovak Republic. For illustration of our method we select life tables of the year 2014.

For Standard Survival model we use estimated parameters from Špirková et al. (2016) presented in Table 1 with respect to expression (2) for three models estimated on male, female and mixed unisex.

Table 1: Estimated constants of the Gompertz' Survival model using gender and unisex life tables of 2014

	$\hat{A}$	$\hat{B}$	$\hat{c}$
Male	0.000000	0.0000689	1.094054
Female	0.000000	0.00000434	1.126396
Unisex	0.001433	0.00001293	1.113202

Source: Špirková et al. (2016)

From Table 1 we can see roughlly "same" parameters $\hat{A}$ and $\hat{B}$ in gender models except for $\hat{c}$ which represents trend drift. As we know unisex tables are weighted mixtures of gender tables what intuitivelly indicates that parameters in unisex model should be also roughlly similar, but they are not.

The Local polynomial regression method uses robustness ideas from the linear regression combined with the local fitting ideas from kernel methods, which involves needed increasing variance in the curve tails. For modelling of survival probability, with respect to expression (3) we use `R` function `loess` in core library developed by Cleveland et al., (1993) with our configuration. Firstly, based on knowledge, we have to fill `R` function only with trimmed data from an age of interest to a point of maxim m_x which represents surface of plausible ages. Secondly, we keep default configurations of kernel function, kernel width parameter and others, except for the surface control set to allow extrapolation. After training algorithm with these configurations we get expected results. Moreover, using presented cookbook is sufficient for fitting any force of mortality function.

Furthermore, estimations of parameters of the Local polynomial Survival model with respect to expression (3) includes $3*$ (plausible ages) parameters and could not be shown in table as Standard Survival model. Therefore, we present the estimation as fitted points represented by line shown in Figure 1. Our intention is to show differences among fitting Standard Survival and non-parametric model namely Local polynomial Survival model. For demonstration we use unisex life tables of the year 2014 see Figure 1.

From plot we can see small difference in curves except for the right tail where the difference is evident. Here is a demonstration of how Local polynomial regression holds whole curve robust average from left to right or otherwise. Based on these curve estimations we present corresponding point estimations with respect to expressions (2) and (3) listed in Table 2 where we add colmumn of differencies. Let us denote $\hat{\mu}_x$ via Gompertz as $\hat{\mu}_x^{(G)}$ and $\hat{\mu}_x$ via Local polynomial as $\hat{\mu}_x^{(LP)}$.

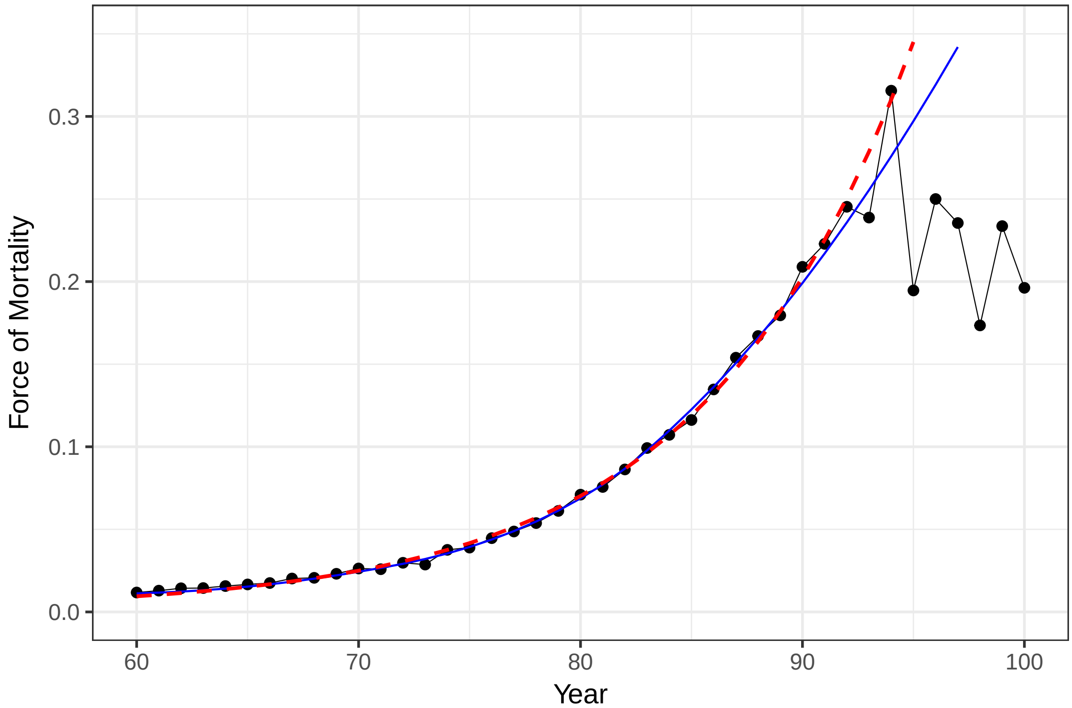
Table 2: Spreadsheet results for point estimations of both Survival models using unisex life tables of 2014

Age	$\hat{\mu}_x^{(G)}$	$\hat{\mu}_x^{(LP)}$	Difference
62	0.01231663	0.01141426	0.00090237
63	0.01312603	0.01254415	0.00058188
64	0.01412610	0.01380196	0.00032414
65	0.01532411	0.01520215	0.00012196
66	0.01672538	0.01676084	-0.00003546
.	.	.	.
.	.	.	.
.	.	.	.
98	0.36562411	0.47550943	-0.10988532
99	0.38999152	0.52917583	-0.13918430
100+	0.41511601	0.58891737	-0.17380136

Source: the author’s work

These estimations are crucial and from the table we can not see average difference or a measure of difference. For this purpose Mean Absolute Error which is a quantity used to mea-

Figure 1: The Force of Mortality calculated from unisex life table represented by points following line, Gompertz Survival model by dashed line and Local polynomial Survival model by solid line



Source: the author's work using library `ggplot2`.

sure how close one estimation is to the eventual estimation. The average difference equals to 0.01912636. In the next section we mention how this small difference influences the amounts of selected product.

4. Comparing on selected product

We continue with the comparision of these models on selected product of the supplementary pension saving, on the basis of which pension annuities can be paid out. The mentioned product includes a permanent monthly annuity and a programmed withdrawal from an accumulated sum at the beginning of retirement time (does not include survivor's benefits) namely Gross monthly pension annuity (*GMA*). Product is given by formula:

$$GMA = \frac{S.\left(1 - \frac{p}{100} - \frac{\alpha}{100}\right)}{m.\ddot{a}_x^{(m)}.\left(1 + \frac{\beta}{1000} + \frac{\delta}{1000}\right)} \tag{6}$$

where

$$\ddot{a}_x^{(m)} = \frac{1}{m}.\sum_{r=0}^{m.(\omega-x-1)+(m-1)} {}_{\frac{r}{m}}p_x.\nu^{\frac{r}{m}}, \tag{7}$$

additional actuary notation:

- S - accumulated sum, gross single premium,
- p - programed withdrawal as % from an accumulated value at the beginning of retirement time,
- i - technical interest rate,
- $v = \frac{1}{1+i}$ - discrete discounting factor,
- x - retirement age,
- m - number of paid, or paid out annuities within one year,
- α - initial costs as a % from accumulated sum,
- β - administrative costs as a ‰ from yearly regular annuity,
- δ - collection costs as a ‰ from yearly regular annuity.

In Tables 3 and 4 we present our solutions of monthly paid out annuities with respect to (6) based on unisex life tables of the year 2014. All monthly paid out annuities are calculated with a basic accumulated sum of 10,000 €. In the product, we use initial costs in an amount of 3% from the accumulated sum, administrative costs in an amount of 3 ‰ from yearly annuity and collection costs in an amount of 1 ‰ form the yearly annuity. In both models we use technical interest rate of 1.9% p.a. in discouting factor for the probability of survival following Špirková et al. (2016).

Presented tables include our offers of monthly paid annuities based on estimations of Survival models with respect to expressions (2) and (3). Moreover, we use additional information about standard deviation in every target age from Local polynomial Survival model. This information makes a possibility to calculate "confidence bounds" of gross monthly pension annuities which can represent a measure of uncertainty of calculated offers. Note that we use default (1-0.05/2) percent quantile of gaussian probability distribution without future information about data generating process. "Confidence bounds" and differences among compared models are also presented in Table 3.

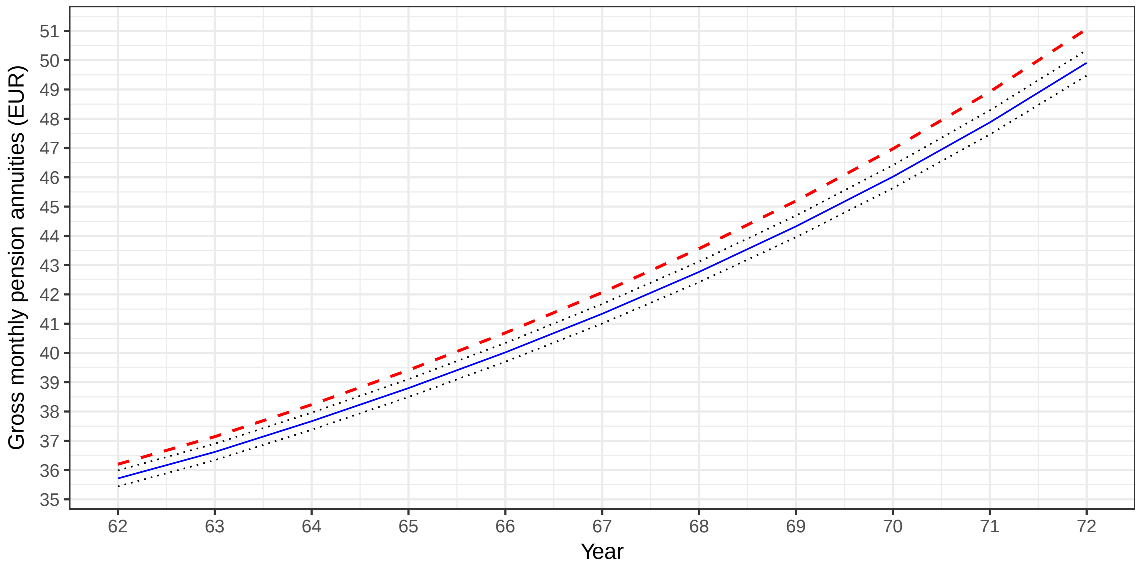
Table 3: The amount of gross monthly pension annuities (€) according to retirement age with an accumulated sum 10,000 €

Retirement age	$GMA^{(G)}$ (€)	$GMA^{(LP)}$ (€)	$GMA^{(LP)}$ left bound (€)	$GMA^{(LP)}$ right bound (€)	Difference (€)
62	36.20	35.71	35.44	35.99	0.49
63	37.14	36.62	36.34	36.90	0.52
64	38.23	37.67	37.38	37.96	0.57
65	39.41	38.80	38.49	39.10	0.61
66	40.69	40.02	39.70	40.34	0.67
67	42.07	41.34	41.01	41.67	0.73
68	43.56	42.77	42.42	43.12	0.79
69	45.19	44.32	43.96	44.69	0.87
70	46.97	46.02	45.63	46.41	0.95
71	48.92	47.87	47.46	48.29	1.05
72	51.06	49.91	49.47	50.35	1.15

Source: the author's work using R

From Table 3, we can see that the impact of the Local polynomial Survival model compared to the Standard Survival model causes an average 1.78% (0.76 €) regress of monthly annuities in 2014 and a difference arising in retirement age from 1.35% to 2.25%. Note that an average difference among estimations listed in analysis is approximatly 0.02 in proportion to 0.76 €. Monthly annuities with respect to the Standard Survival model are also out of "confidence bounds" which indicates strong difference among used models. Moreover, we can see a reduction in monthly pension incomes from pension annuities with the Local polynomial Survival model ilustrated in Figure 2.

Figure 2: Gross monthly pension annuities calculated from the Local polynomial Survival model following solid line, GMA via Standard Survival model by dashed line and "Local polynomial Survival model confidence bounds" by dotted lines



Source: the author's work using library `ggplot2`.

Figure clearly shows annuities results and gives us a strong intuition about intersection of both models with decresing years, on the other side is inverse impact. Both annuity curves sligthly mimic their force of mortality estimations, see Figure 1. Mentioned regress of monthly annuity we can be understood as "a price" which individual pays for a higher survival probability in ages between 90-100 from the Local polynomial Survival model.

In the last part, we use technical interest rates 0.7% ,1.9% and 2.5% to compare the impact of the technical interest rate on the amount of the pension annuity. In calculations we apply same process as we present in Table 3 for each technical rate. Technical interest rates also have a significant impact on the amount of monthly annuities. The increase of the technical interest rate from 0.7% to 2.5% causes the decrease of average differences among mentioned models from $2,13\%$ to $1,64\%$, for more details see Table 4.

Table 4: Average amounts of gross monthly pension annuities (€) according to the technical interest rate with an accumulated sum 10,000 €

Technical rate	$GMA^{(G)}$	$GMA^{(LP)}$	$GMA^{(LP)}$ left bound	$GMA^{(LP)}$ right bound	Average difference
$i = 0.7$	36.64	35.86	35.55	36.17	0.78
$i = 1.9$	42.68	41.91	41.57	42.26	0.76
$i = 2.5$	45.86	45.11	44.75	45.47	0.75

Source: the author's work using R

5. Conclusion

In this paper we discussed modelling the force of mortality using a new non-parametric approach with comparision to the Standard Survival model based on Gompenz law. In particular, we focused on the impact of modelling based on selected product offers of monthly annuities of the year 2014 with respect to unisex life tables. Moreover, we used "confidence bounds" which could give us an unknown uncertainty of our offers.

Suggested improvements of modelling make survival probabilities higher for older people which is fairer to them, however this leads to lower offers of monthly annuities. It would be interesting to use this approach for products targeted to customers with at age below 80.

In the further research, we plan to study the impact of "the force of mortality confidence bounds" from the more complex statistical point of view using life tables from last years. That can provide extra knowledge about a data generating process or a probability distribution. Furthermore, we will hold the information from last years as an inference for this or next year. Therefore, we would like to improve and extend presented modelling of "confidence bounds" to cover uncertainty. This bounds can be used to view the future of the force of mortality and prices of selected products.

Acknowledgements

Author has been supported by the Slovak Scientific Grant Agency VEGA NO. 1/0093/17 - Identification of risk factors and their impact on products of the insurance and saving schemes.

References

[1] Hastie, T., Tibshirani, R., Friedman, J. 2009. The Elements of Statistical Learning. USA: Springer Science+Business Media. ISBN 978-0-387-84857-0

[2] Dickson, D. C. M. et al. 2009. Actuarial mathematics for life contingent risks. New York : Cambridge University Press, 2009.

[3] Špirková J., Kollár I., Spišiaková M. 2016. Notes to the product of the supplementary pension saving scheme. AMSE 2016

[4] Cleveland , W. S. , Devlin , S. J. 1988. Locally weighted regression: An approach to regression analysis by local fitting . J. Am. Statist. Assoc. 83 : 596 – 610.

[5] Wickham, H. 2016 ggplot2, Elegant Graphics for Data Analysis (Second Edition) USA: Springer Use R. ISBN 978-3-319-24275-0

[6] R Core Team 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, `http://www.r-project.org/`.

[7] National Bank of Slovakia 2013. The Order of the National Bank of Slovakia of 25 June 2013. [Vestník NBS z 25. júna 2013]. [cit.17-09-2013] `http://www.nbs.sk/_img/Documents/_Legislativa/_Vestnik/ciastka23V-2013.pdf`.

[8] National Bank of Slovakia 2015. The Order of the National Bank of Slovakia of 1 December 2015. [Vestník NBS opatrenie NBS č. 25/2015]. [cit. 11-03-2016] `http://www.nbs.sk/_img/Documents/_Legislativa/_Vestnik/OPAT25-2015.pdf`.

[9] Faraway, J. J. 2016 Extending the linear model with R (second edition): Chapman and Hall/CRC. ISBN 978-1-498-72096-0

[10] Gompertz, B. 1825 On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical transactions of the Royal Society of London, 115, 513-583.

[11] Makeham, W. M. 1860 On the law of mortality and construction of annuity tables. The Assurance Magazine and Journal of the Institute of Actuaries, 8(06), 301-310.

[12] Ledolter, J. 2008 Smoothing Time Series with Local Polynomial Regression on Time. Communications in Statistics - Theory and Methods, 37, 959-971.