

# EFFECTIVE USE OF MS EXCEL IN EDUCATION OF ECONOMIC DEMOGRAPHERS

*Ondrej Simpach*

ondrej.simpach@vse.cz

Faculty of Informatics and Statistics,  
University of Economics in Prague

---

## **ABSTRACT**

This study presents the possibilities of using optimization tools in an office package MS Excel (various versions) for specific purposes in education of economic and actuarial demographers, because these purposes are necessary in advanced higher education courses. For special types of applications it is important to estimate and refine the parameters of specific functions. One of these function used in demography is Gompertz–Makeham function for balancing specific death rates. The solver in MS Excel 2007 was equipped with only basic module for solving linear programming and simple nonlinear programming. With the coming of version of MS Excel 2010 the modules GRG–Nonlinear and Evolutionary have been incorporated into the solver, which can be used to solve more difficult nonlinear problems. These modules are available from the 2010 release in the basic installation of the base price, so there is not necessary to re-buy it externally. The aim of this study is to show the possibilities of levelling specific death rates in MS Excel solver and compare the computational times that are needed to obtain the estimates.

## **KEYWORDS**

MS Excel 2010, solver, economic demography, specific death rates, optimization

## **INTRODUCTION**

In basic applications and courses of balancing specific death rates, which are important in life insurance, we use moving averages of different lengths and moving averages of different set of weights. Given that the weighted moving averages are not enough, it is useful to choose one of the sophisticated functions. One of these functions, which is taught in advanced economic and actuarial demographic courses, is Gompertz–Makeham and also modified Gompertz–Makeham function, (see Gompertz, (1825), and Makeham, (1860)), which are both of exponential type. The initial estimates of parameters can be set up manually and using by other software it is possible to refine it. Given that the optimization software is relatively expensive both for education or home use; we can use the MS Excel Solver, which is able to solve simple linear programming (LP) and non-linear programming (NLP) tasks within a reasonable time. Even in version 2007 there was available basic method to solve simple LP and NLP only, and the additional package was necessary to purchase externally. Since version 2010 is incorporated in standard package GRG–Nonlinear and Evolutionary method, which solve more difficult nonlinear problems in less time than before. A subsequent study will provide a view on the education of economic and actuarial demographers using effective levelling of specific death rates in relatively cheap software and the comparison of computational time needed to obtain accurate estimates of parameters of functions in solvers version 2007 and 2010, based on real data of Czech Republic and other EU countries. At the end we compare these

computational times with the selected professional system and compare the advantages and disadvantages in effectiveness of the use in education.

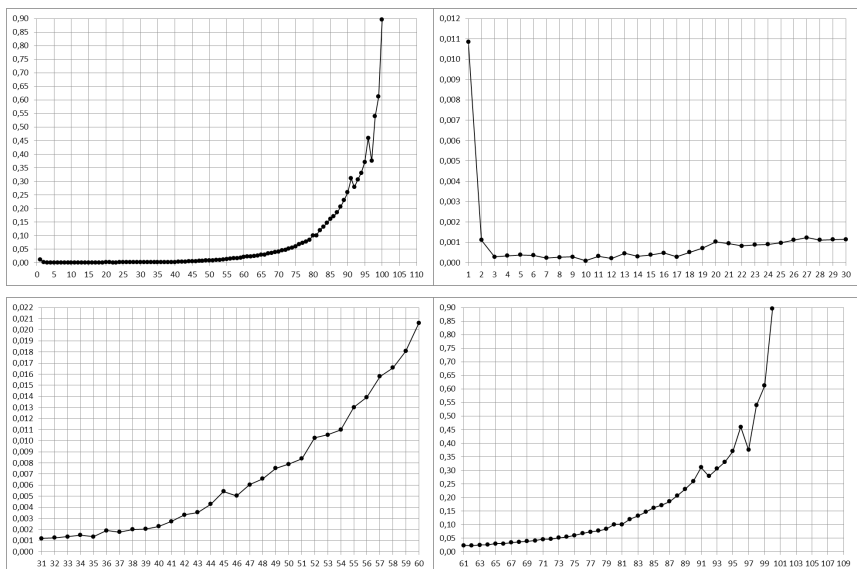
## **THE POSSIBILITIES OF OPTIMIZATION**

Optimization approaches are important in economic disciplines and in demography as well, because for special types of applications it is necessary to estimate and refine the parameters of specific functions. One of these function used in demography is Gompertz-Makeham function for balancing specific death rates, which is used in the construction of complete life tables for commercial use (see e.g. the projection of the human capital of the Czech Republic and its regions to 2050 by Fiala et al., (2011), or the development of life expectancy in the Czech Republic and EU countries by Langhamrová Jana and Langhamrová Jitka, (2010)). This function is of exponential type and therefore is effective use a special procedure for its solution. The initial estimates of its parameters can be calculated from empirical data (Cipra, (2003)) and using the optimization software is also possible to improve these estimates. Because the optimization software is expensive and in comparison with other applications of operations research, the balancing of specific death rates is relatively simple application, it is not effective to invest for buying professional solver, but it is better to use available solutions in a readily available application, e.g. package MS Office.

The solver in MS Excel 2007 was equipped with only basic module for solving LP and simple NLP. Additional modules for solving more difficult problems were only available after the external purchase. This has raised the question of whether is better to buy optimization software directly than a supplement to Microsoft Office suite. With the coming of version of MS Excel 2010 the modules GRG–Nonlinear and Evolutionary have been incorporated into the solver, which can be used to solve more difficult NLP. These modules are available from the 2010 release in the basic installation of the base price, so there is not necessary to re-buy it externally. The solution with packages and solvers of MS Office is already at well advanced level. It is highly likely that with the advent of new version of MS Office 2013 and new versions of solvers, the possibilities will be much more varied and more efficient.

## **MATERIAL AND METHODS**

In advanced courses of education of economic and actuarial demographers we use the modelling of specific death rates particularly for the further purpose of life insurance. The life tables describe the dependence of mortality on the age. The basic characteristics for their calculation are therefore specific death rates. The development of specific death rates for the case of Czech Republic in 2005 is shown for illustration in Fig. 1.



**Fig. 1: The specific death rates for the case of the Czech Republic in 2005 (total population, empirical data). Source: CZSO, own construction**

In the case of life tables for one calendar year  $t$  (see e.g. Jacobs et al. (1999)) are specific death rates calculated by the formula

$$m_x = \frac{M_{t,x}}{\frac{S_{t,x} + S_{t+1,x}}{2}}, \quad (1)$$

where  $M_{t,x}$  is the number of death in the age of  $x$  in year  $t$ ,  $S_{t,x}$  is the initial state of  $x$ -years old person in year  $t$ .

Death rates are developed up to particular age  $x$  (usually  $x = 60$ ) by previously expected system. For different types of populations are these values more or less scattered in higher age, and for the purpose of further analysis is necessary to balance them. This balancing is performed using either a weighted moving averages (simple method), or with sophisticated balancing function (complex method). In the past it was common use for balance of specific death rates the simple moving averages. Moving averages are used with a length of three,

$$\tilde{m}_x^{(3)} = \frac{m_{x-1} + m_x + m_{x+1}}{3} \quad \text{for } x \in \langle 3; 5 \rangle, \quad (2)$$

respectively with a length of nine,

$$\tilde{m}_x^{(9)} = 0,2 \cdot m_x + 0,16 \cdot (m_{x-1} + m_{x+1}) + 0,12 \cdot (m_{x-2} + m_{x+2}) + 0,08 \cdot (m_{x-3} + m_{x+3}) + 0,04 \cdot (m_{x-4} + m_{x+4}) \quad \text{for } x \in \langle 6; 29 \rangle, \quad (3)$$

respectively with a length of nineteen values

$$\begin{aligned} \tilde{m}_x^{(9)} = & 0,2 \cdot m_x + 0,1824 \cdot (m_{x-1} + m_{x+1}) + 0,1392 \cdot (m_{x-2} + m_{x+2}) + \\ & + 0,0848 \cdot (m_{x-3} + m_{x+3}) + 0,0336 \cdot (m_{x-4} + m_{x+4}) - 0,0128 \cdot (m_{x-6} + m_{x+6}) - \\ & - 0,0144 \cdot (m_{x-7} + m_{x+7}) - 0,0096 \cdot (m_{x-8} + m_{x+8}) - 0,0032 \cdot (m_{x-9} + m_{x+9}) \end{aligned}$$

for  $x \in \langle 30; 59 \rangle$ , (4)

as used for example Dotlačilová et al. (2012). Since higher age than 60 years can no longer be used. High quality outputs provides Gompertz–Makeham balancing function for specific death rates (see e.g. Jodrá, (2009), or Gavrilova, Gavrilov (2011)), defined as

$$\tilde{m}_x^{(GM)} = a + b \cdot c^{x+\frac{1}{2}} \quad (5)$$

for which we set up the beginning of the first age interval  $x_0 = 60$  and the length of interval  $k = 8$ . The initial estimations of parameters of Gompertz–Makeham function are set according to the recommendations by the following algorithm, (see Fiala (1999)):

Calculate the empirical sums of specific death rates  $G_1$ ,  $G_2$  and  $G_3$  in the various intervals as

$$G_1 = \sum_{x=60}^{67} m_x, \quad G_2 = \sum_{x=68}^{75} m_x \quad \text{and} \quad G_3 = \sum_{x=76}^{83} m_x, \quad (6)$$

where  $x = 60, 61, \dots, 67$  completed years of life for the first sum, respectively  $x = 68, 69, \dots, 75$  completed years of life for the second sum and  $x = 76, 77, \dots, 83$  completed years of life for the third sum. Next calculate  $c^k$  and  $c$  (where  $k = 8$ ) as

$$c^8 = \frac{G_3 - G_2}{G_2 - G_1} \quad \text{and} \quad c = \sqrt[8]{c^8}, \quad (7)$$

Finally we are able to calculate the value of the auxiliary expression

$$K_c = c^{60,5} \cdot (1 + c + c^2 + \dots + c^7) = c^{60,5} \cdot \frac{c^8 - 1}{c - 1} \quad (8)$$

and then calculate the last unknown parameters of Gompertz–Makeham function  $b$  and  $a$  as

$$b = \frac{G_2 - G_1}{K_c \cdot (c^8 - 1)} \quad \text{and} \quad a = \frac{G_1 - b \cdot K_c}{8} \quad (9)$$

The values of Gompertz–Makeham function (5) we calculated for the values of age from 60 to 83 years. As the values of the parameters  $a$ ,  $b$  and  $c$  we use the values, that will refine the least squares method. For the age interval  $x \in \langle 83; 110 \rangle$  we use instead of the original Gompertz–Makeham function its modification defined as

$$\tilde{m}_x^{(GM)'} = a + b \cdot c^{8 + \lceil \ln(d \cdot (x + \frac{1}{2} - 8) + 1) \rceil / d} \quad (10)$$

whereas an initial estimate of the parameter  $d$  we choose for example  $d = 0.02$  (see recommendations by e.g. Fiala (1999)). Next, for age 83 we overwrite the value given by Gompertz–Makeham function (5) with modified Gompertz–Makeham function given

by formula (10). The value of  $d$  will be optimized using least squares method. Weighted squares of deviations, which are defined as

$$WSD = \frac{S_{t,x} + S_{t+1,x}}{2 \cdot m_x \cdot (1 - m_x)} \cdot (m_x - \tilde{m}_x^{(GM)})^2 \quad (11)$$

we have to calculate for all values of  $x$  from age 60 to  $y$ , where  $y$  is the maximum age for which we have empirical specific death rates with non-zero value. Weighted sums of squared deviations we calculate separately for the age in range 60–82 years (the sum of  $S_1$ ), where we use the original (unmodified) Gompertz–Makeham function, and particularly for the age in range of 83– $y$  (sum of  $S_2$ ), where we use a modified function. Optimization procedure is performed with the “Solver”, implemented in MS Excel. Optimization is performed twice: first separately for sum of squares  $S_1$  (when we change the parameters  $a$ ,  $b$  and  $c$ ) and separately for the second sum of squares  $S_2$  when we change only the parameter  $d$ . Because we used the empirical values of specific death rates to estimate the weights, we do not have to insert any limiting conditions.

## RESULTS AND DISCUSSION

Using the algorithm described above there were performed the initial calculations of unknown parameters (see Tab. 1) and the initial calculations of parameters of Gompertz–Makeham function, that were subsequently improved by the optimizing process. These initial values are shown in Tab. 2 for the case of the Czech Republic in 2005.

$G_1$	$G_2$	$G_3$	$c^8$	$K_c$
0,228681	0,406511	0,813220	2,287064745	6157,982822

**Tab. 1: The initial estimates of unknown parameters. Source: own construction**

$a$	$b$	$c$	$d$	$S_1$	$S_2$
0,011314	0,000022	1,10894449	0,020000	43,17243991	42,08434903

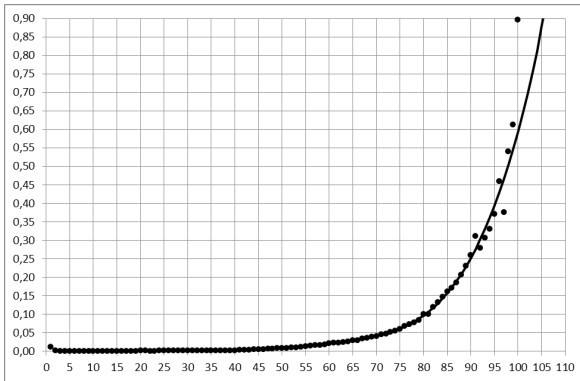
**Tab. 2: The parameters of Gompertz–Makeham function, the initial estimates of unknown parameters – The Case of Czech Republic in 2005. Source: own construction**

Using the optimization solver (any version - a decisive was only the time needed to achieve the result), the initial values of parameters were refined. This improvement in the case of the Czech Republic in the year 2005 is shown in Tab. 3.

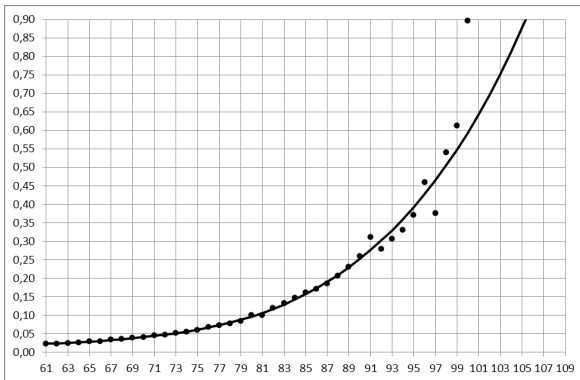
$a$	$b$	$c$	$d$	$S_1$	$S_2$
0,012019	0,000015	1,11486900	0,020000	38,4745460	40,0177650

**Tab. 3: The parameters of Gompertz–Makeham function, the estimates refined by least squares method – The Case of Czech Republic in 2005. Source: own construction**

The improved parameters of Gompertz–Makeham functions were used to balance of specific death rates - to illustrate the case of the Czech Republic in 2005 is shown in Fig. 2, the detail (from age 60 to age  $\omega$ ) is shown in Fig. 3.

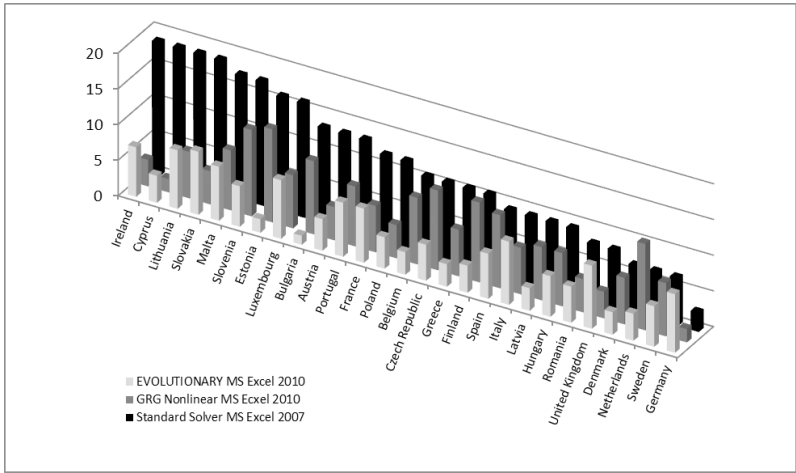


**Fig. 2: The balanced specific death rates (the line – by Gompertz–Makeham and modified Gompertz–Makeham function) for the case of the Czech Republic in 2005 (total population). Source: CZSO, own construction**



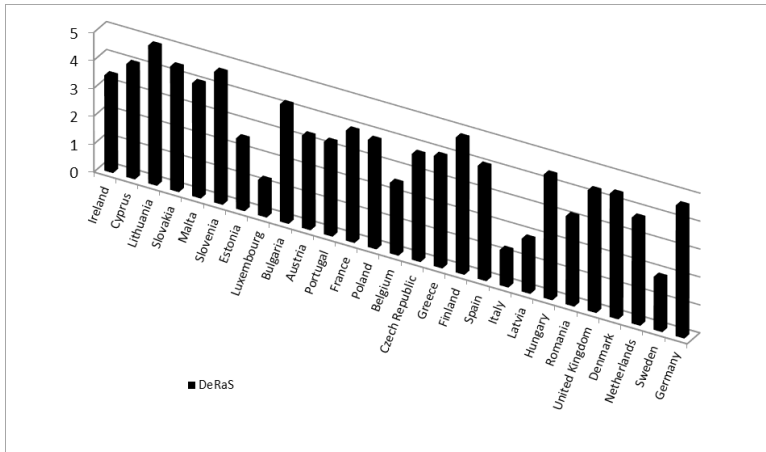
**Fig. 3: The balanced specific death rates (the line, detail view from 61 to  $\omega$  completed years – by Gompertz–Makeham and modified Gompertz–Makeham function) for the case of the Czech Republic in 2005 (total population). Source: CZSO, own construction**

The procedure of estimating unknown parameters (written above, see chapter Material and Methods) was repeated ten times for Czech Republic and other selected EU countries with the concrete solver and the computing times were measured. Ten times the measurements were taken in a standard module in solver of MS Excel 2007, ten times the measurements were taken in the module GRG Nonlinear in solver in MS Excel 2010, and finally ten times the measurements were taken in the module Evolutionary in solver in MS Excel 2010. From these times there was calculated the average time, that has been necessary for the calculation of the corrected estimations of parameters of Gompertz–Makeham function and modified Gompertz–Makeham function. The used CPU was Intel® Core™ i5-2430 2.40 GHz running at 64-bit platform. An overview of these computational times is summarized in Fig. 4 in seconds.



**Fig. 4: The averages of computational times (in seconds) in the case of the Czech Republic and other selected EU countries in 2005. Source: own calculation and construction**

For comparison of computational times between solvers of MS Excel with professional system DeRaS (Burcin et al. (2012)), which is developed for levelling specific mortality rates, we can use Fig. 5, which displays the results of this calculations. The individual times are in some cases comparable to MS Excel Evolutionary solver, in some cases are lower.



**Fig. 5: The averages of computational times (in seconds) in the case of use professional system DeRaS (Burcin et al. (2012)). Source: own calculation and construction**

## CONCLUSION

It is clear that almost in all cases the calculation module GRG Nonlinear provides solutions in less computing time than the standard module, implemented in MS Excel 2007. In most cases, the Evolutionary module provided an optimal solution before GRG Nonlinear

module, but it is not the case. The solution provided by individual modules, of course, depends on the input data and on other circumstances. However, we can confirm, that the modules GRG Nonlinear and Evolutionary, implemented in MS Excel 2010, provide a solution sooner than in older versions of MS Excel, what can bring benefits for specific groups of users and analysts. This study presented the possibilities of using optimization tools in an office package MS Excel (versions 2007 and 2010) for specific purposes in education of economic and actuarial demographers, because these purposes are necessary in advanced higher education courses. In the case of using a professional system, almost all estimates are obtained in less time than using MS Excel Solver. The differences, however, have a low significance. We hope that with the advent of new version of MS Office 2013 and new versions of solvers, the possibilities will be much more varied and more efficient.

## REFERENCES

- Burcin, B., Hulíková Tesárková, K., Kománek, D. (2012). DeRaS: software tool for modelling mortality intensities and life table construction. Charles University in Prague, Prague. <http://deras.natur.cuni.cz>
- Cipra, T. (2003). Life tables for Czech pension funds. *Insurance: "Mathematics & Economics"* **32** (2003), 176–177.
- Dotlačilová, P., Langhamrová, J., Šimpach, O. (2012). Vybrané logistické modely používané pro vyrovnávání a extrapolaci křivky úmrtnosti a jejich aplikace na populace vybraných zemí Evropské unie. *Forum Statisticum Slovacum [online]*, 2012, roč. 8, č. 7, s. 21–25.
- Fiala, T. (1999). Mortality in the Age from 20 to 70 Years in the Czech Republic in the Period 1950 to 1996. *Acta Oeconomica Pragensia*, 1999, roč. 7, č. 4, s. 35–44.
- Fiala, T., Langhamrová, J., Průša, L. (2011). Projection of the Human Capital of the Czech Republic and its Regions to 2050. *Demografie*, 2011, roč. 53, č. 4, s. 304–320.
- Gavrilova N. S., Gavrilov L. A. (2011). Ageing and Longevity: Mortality Laws and Mortality Forecasts for Ageing Populations [In Czech: Stárnutí a dlouhověkost: Zákony a prognózy úmrtnosti pro stárnoucí populace]. *Demografie*, **53** (2): 109-128.
- 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** (1825): 513–585.
- Jacobs, D. R. Jr, Meyer, K. A., Kushi, L. H. and Folsom, A. R. (1999). Is whole grain intake associated with reduced total and cause-specific death rates in older women? The Iowa Women's Health Study. *American Journal of Public Health*, March 1999: Vol. 89, No. 3, pp. 322-329.
- Jodrá, P. (2009). A closed-form expression for the quantile function of the Gompertz-Makeham distribution. *Mathematics and Computers in Simulation* **79** (10): 3069–3075.
- Langhamrová, Jana, Langhamrová, Jitka. (2010). Vývoj střední délky života v České republice a v zemích EU. In: *Bratislava : SAV; ČAV*, 2010. 36–44 s.
- Makeham, W. M. (1860). On the Law of Mortality and the Construction of Annuity Tables. *Journal of the Institute of Actuaries and Assurance*. **8** (1860): 301–310.