An application of Extreme Value Theory for modelling life expectancy

Grażyna Trzpiot[1], Justyna Majewska[2]

Abstract

In this paper we have applied new method (Medford, 2015) in order to project the life expectancy in Poland using life expectancy values from different European countries. This new approach combines concept of extreme value distribution and "best life expectancy" idea of Oeppen and Vaupel (2002), who demonstrated that the development of "best practice" life expectancy is best approximated by a linear trend, estimated over the past 160 years. Approach based on EVT has never been applied to life expectancy projections in Poland. The method produces results (life expectancy at birth and age 65) that are similar to those obtained by Central Statistical Office of Poland for period 2015-2050.

Keywords: *EVT, extreme value distribution, life expectancy*

JEL Classification: C130, C180

1 Introduction

Life expectancy at birth is defined as the mean number of years still to be lived by a person at birth, if subjected throughout the rest of his or her life to the current mortality conditions. Over the last 170-year period European countries have experienced a continuation of the pattern of falling mortality rates that began in the 19th century. Over long historical period we have observed linear trend in life expectancy. However~~,~~ today there is no consensus on trends over the very long term (European Commission 2014). Official projections generally assume that gains in future life expectancy at birth will slow down compared with historical trends.

Accuracy is crucial in a life expectancy projection. It allows governments and other institutions to plan wisely and helps individuals comprehend the likely futures for their countries and the world.

As Li and Lee (2005) indicated the convergence in mortality levels for closely related populations can lead to unsuitable mortality projections, if the projections for individual populations are obtained in isolation from one another. Similar historical trends in long-run life expectancy patterns can be useful for countries. Besides, analyses of the main determinants of life expectancy (the socio-economic, environmental or behavioural factors) of associated populations are crucial. Knowledge of existence of some common stochastic trends

[1] Corresponding author: University of Economics in Katowice, Department of Demography and Economic Statistics, 1 Maja 50, 40-287 Katowice, e-mail: justyna.majewska@ue.katowice.pl.

[2] University of Economics in Katowice, Department of Demography and Economic Statistics, 1 Maja 50, 40-287 Katowice, e-mail: grazyna.trzpiot@ue.katowice.pl.

in mortality rate in cluster of European countries (see Lazar et al. 2016) can be used for projections of mortality rates and life expectancy.

In this paper we have applied new method (Medford, 2015) in order to project the life expectancy in Poland using life expectancy values from different European countries. This new approach combines concept of extreme value distribution and “best practice“ life expectancy idea of Oeppen and Vaupel (2002). “Best practice” life expectancy reffers to the maximum life expectancy observed among national populations at a given age, and according to Oeppen and Vaupel (2002), the development of “best practice” life expectancy is best approximated by a linear trend, estimated over the past 160 years. The approach based on EVT has never been applied to life expectancy projections in Poland.

2 Basic of Extreme Value Theory

The EVT has been proved to be a powerful tool to study extreme event distributions and widely used in many applications in multidisciplinary areas. EVT is well documented in the literature (e.g. Coles, 20012; Gilli and Këllezi, 2006; Embrechts et al. at al, 1997).

The foundation of extreme value theory is a family of probability distributions known as extreme value distributions. For $\xi \neq 0$ the generalized extreme value GEV distribution function is given by

$$P(Y \leq y) = G(y;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{1}$$

where Y is some random variable and the three parameters $\mu(-\infty < \mu < \infty)$, $\sigma(\sigma > 0)$ and $\xi(-\infty < \xi < \infty)$ are referred to as the location, scale and shape parameters, respectively. The location parameter indicates the center of the distribution, the scale parameter indicates the size of deviations around the location parameter, and the shape parameter governs the tail behavior of the GEV distribution. All three parameters are important, but ξ is especially critical for extrapolation of extreme events. The shape ξ parameter determines the heaviness of the right tail and this leads to three types of distributions. When $\xi < 0$, the distribution has a bounded upper finite end point and is short-tailed and in this case the distribution is often called a Weibull distribution. When $\xi > 0$, there is polynomial tail decay leading to heavy tails and the GEV distribution is called then Fréchet. The case where ξ=0 is taken to be the limit of eq. (1) as $\xi \rightarrow \infty$, and there is exponential tail decay leading to light tails (usually referred to as the Gumbel type).

Generally there are two ways of identifying extremes in data: block maxima and peak over threshold method. Block maxima method considers the maximum the variable takes in successive periods. These selected observations constitute the extreme events, also called block maxima.

The limit law for the block maxima, which we denote by M_n, with n the size of the subsample (block), is given by the Extremal Types Theorem (Fisher and Tippett, 1928; Gnedenko, 1943): if there exists sequences of constants $\{a_n>0\}$ and $\{b_n\}$, such that as $n \to \infty$

$$P\left(\frac{M_n - b_n}{a_n} \le y\right) \to G(y)$$

where $G(y)$ is a non-degenerate distribution function, then G must be a member of the GEV family of distributions.

In estimating the parameters μ, σ and ξ from observations, there are a number of popular methods including maximum likelihood estimation (MLE), Bayesian methods, and the L-moments technique.

Estimates of extreme quantiles of the maxima are obtained by solving for y in eq. (1):

$$y_p = \mu - \frac{\sigma}{\xi}\left[1 - \left[-\log(1-p)\right]^{-\xi}\right] \tag{2}$$

where the distribution function of the GEV, $G(y_p)=1-p$, and p is the tail probability or the probability of realising a value at least as large as y_p. When we focus on annual maxima then on average the quantile y_p is expected to be exceeded with probability p or on average once every $1/p$ years (Coles, 2001).

3 Application

Life tables were obtained from Human Mortality Dataset (HMD, 2017)[3]. Calculations were made in R with *spdep, demography, evd* and *segmented* packages[4].

Best life expectancies were calculated for the following European countries: Denmark, Finland, Norway, Sweden, Netherlands, Belgium, France, Italy, Ireland, Portugal, Spain, UK, Italy, Austria, and Czech Republic. Existence of similar historical trends in life expectancy

[3] http://mortality.org.

[4] Package *spdep* – a collection of functions to create spatial weights matrix objects from polygon contiguities, package *demography* – functions for demographic analysis, *demography* - functions for demographic analysis including lifetable calculations, *evd* – provides extends simulation, distribution, quantile and density functions to univariate and multivariate parametric extreme value distributions, *segmented* – for given a regression model, segmented „updates“ the model by adding one or more segmented (i.e., piecewise-linear) relationships.

was motivation for grouping countries into the clusters. SKATER algorithm[5] divided countries into homogeneous contiguous clusters according to the following variables: gross domestic product per capita, educational attainment, fertility and access to health care[6]. In this analysis we skipped cluster that contains most countries from Eastern and Southern Europe. The first cluster groups developed countries with the highest GDP per capita and very high fertility rate. The second cluster includes countries with the highest fertility rate and low education level. The third cluster groups countries with the lowest percentage of population aged 16 and over, that they had unmet needs for medical examinations or treatment. Lazar et al. (2016) proved the existence of some common stochastic trends determining the mortality in European countries.

Countries in cluster	**Historical period**
1. Denmark, Finland, Norway, Sweden, Netherlands	1960-2012
2. Belgium, France, Italy, Ireland, Portugal, Spain, UK	1950-2012
3. Italy, Austria, Czech Republic	1950-2012

Table 1. Clusters of European countries with common mortality trends.

Historical trends in life expectancy are observed in Fig. 2. and Fig. 3. For comparison purpose historical trend of life expectancy in Poland was added. Countries from the first cluster reached very high life expectancy sooner than the rest European countries, however the speed in increase between 1960 and 2010 was not as fast as it was in the countries from the second and the third cluster. Since 1960, in countries from the first cluster life expectancy at birth has typically increased by 6-9 years for female, meanwhile in the second cluster it was 8-12 years. In the third cluster trends are remarkably similar until ‘70s. This parallelism is rarely remarked later.

[5] Spatial Kluster Analysis by Tree Edge Removal is an alternative to other regionalization methods, based on minimum spanning trees (Assunção et al. 2006).
[6] Demographic and economic data (recorded in 2012) used for spatial clustering were obtained from Eurostat database.

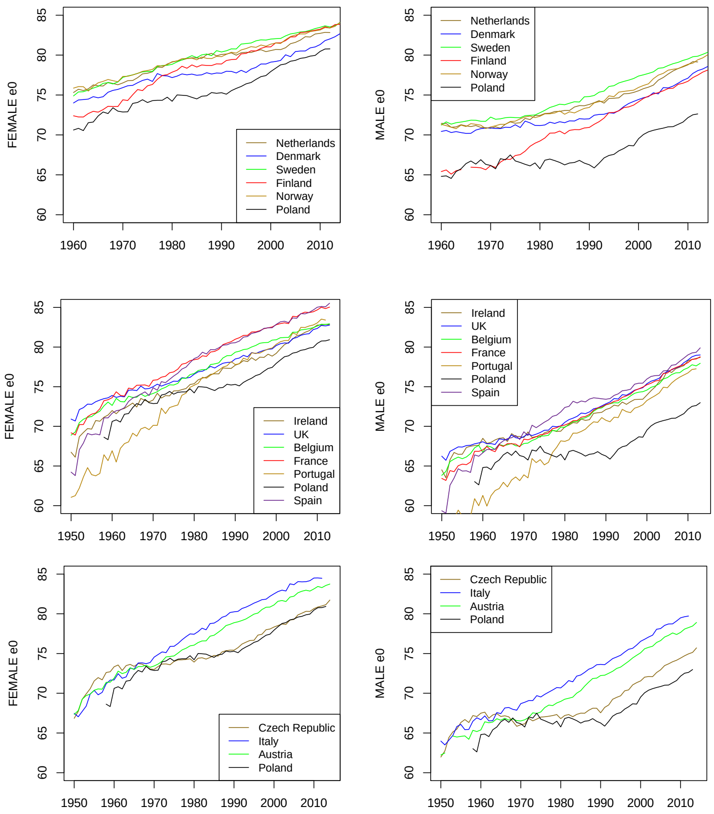

Fig. 1. Historical trends of life expectancy at birth in each cluster.

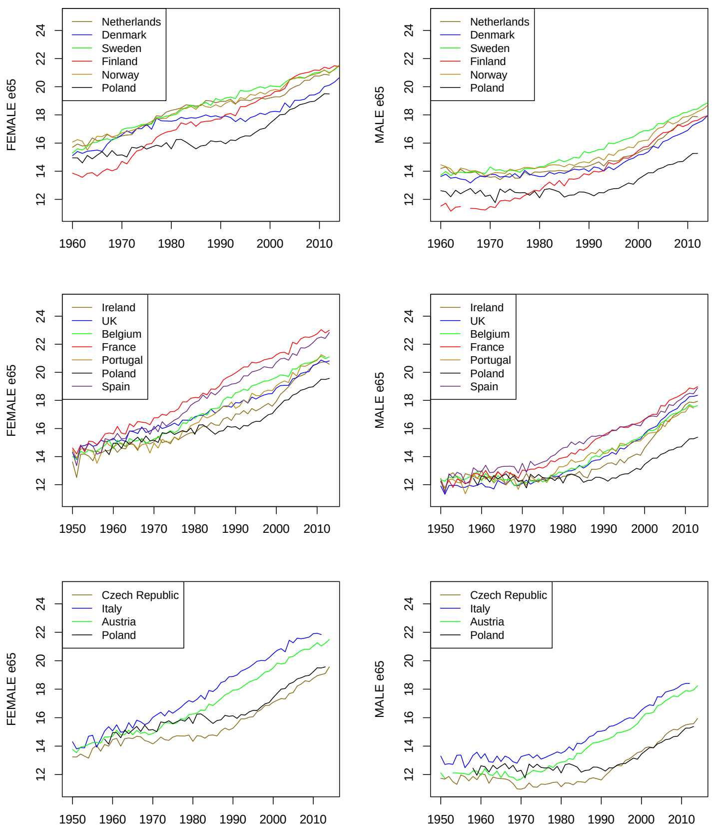

Fig. 2. Historical trends of life expectancy at 65 in each cluster.

Figures 3-5 show countries that has the highest period life expectancy by male and female, at birth[7]. Over much of the periods we observe very strong linear trends over time.

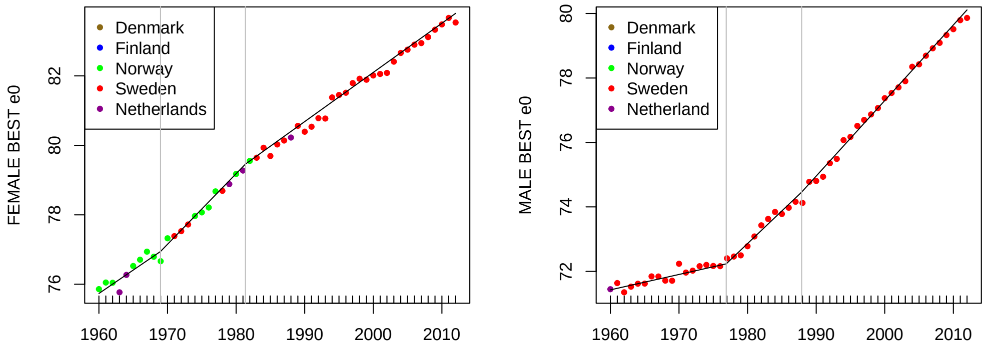

Fig. 3. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 1.

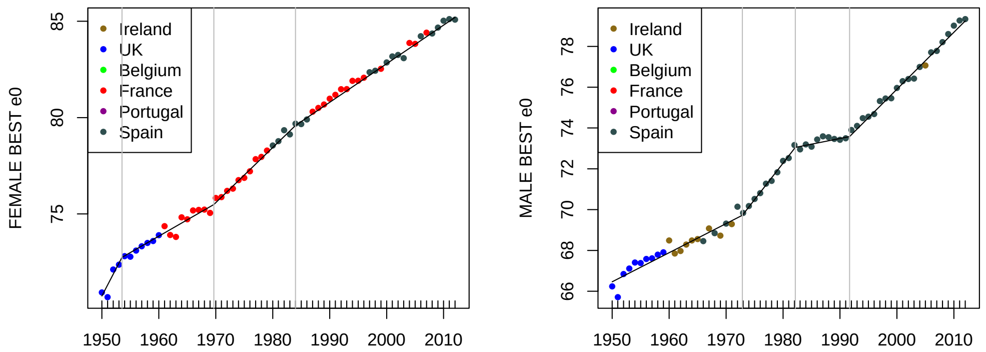

Fig. 4. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 2.

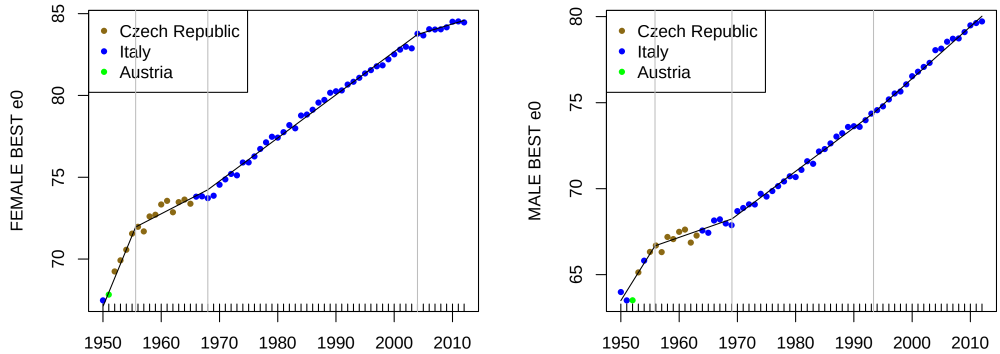

Fig. 5. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 3.

[7] Because the pages limitation, plots for the highest period life expectancy by male and female at age 65 are not presented in this paper.

Particularly interesting is the fact that life expectancy at birth for women has been slowing in the last decade, while life expectancy at birth for men has been accelerating. Acceleration in life expectancy is also observed by authors for people over age 65. In the first step the presence of differential rates of increase was investigated with Davies test (Davies, 2002). In the second step the break points were founded (marked as vertical lines in Fig. 4.-6.). As a result we got different fitted linear regressions between these break points. Followed by Medford (2015) the GEV model was fitted to the data beginning at the most recent break point, thus ensuring that the correct speed of life expectancy increase is captured as accurately as possible.

We fitted a time-dependent GEV model, GEV(μ_t, σ, ξ), where $\mu_t = \beta_0 + \beta_1 t$ is the location parameter, σ is the scale parameter, ξ is the shape parameter, and $t = 1...t_{max}$, where t_{max} is the last calendar year of the data set. Maximum likelihood method was used for parameters estimation. Results are presented in Table 2. – Table 4.

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	72.49 (0.11)	0.26 (0.01)	0.20 (0.03)	-0.07 (0.13)
Female e0	78.81 (0.03)	0.20 (0.00)	0.16 (0.02)	-0.57 (0.11)
Male e65	14.53 (0.05)	0.14 (0.00)	0.09 (0.02)	-0.16 (0.17)
Female e65	18.03 (0.03)	0.10 (0.00)	0.17 (0.05)	-0.80 (0.24)

Table 2. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 1).

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	73.39 (0.06)	0.28 (0.00)	0.18 (0.04)	-0.68 (0.29)
Female e0	80.71 (0.13)	0.11 (0.00)	0.91 (0.04)	-0.73 (0.21)
Male e65	15.30 (0.05)	0.16 (0.00)	0.21 (0.04)	-0.66 (0.30)
Female e65	20.88 (0.13)	0.15 (0.01)	0.20 (0.05)	-0.50 (0.22)

Table 3. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 2).

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	74.30 (0.07)	0.30 (0.01)	0.15 (0.03)	-0.11 (0.19)
Female e0	83.55 (0.12)	0.11 (0.00)	0.13 (0.01)	-0.82 (0.53)
Male e65	17.69 (0.04)	0.14 (0.02)	0.06 (0.01)	-1.03 (0.78)
Female e65	21.48 (0.00)	0.06 (0.00)	0.02 (0.01)	1.20 (0.65)

Table 4. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 3).

Inspired by Medford (2015) projections of life expectancy for Poland were produced. These were done by updating the time-varying location parameter of the model μ_t for future values of time *t*. In order to make projections, the time varying GEV model was fitted and then the 50th percentile calculated (see eq. (2)). Projected values of the median of the fitted GEV distribution were used to approximate future values of life expectancy. Results are presented in Table 5. – Table 7.

Scenario 1	**2015***	**2025**	**2030**	**2035**	**2040**
Male e0	73.60	76.10	77.40	78.70	80.00
Female e0	81.60	81.55	82.55	83.55	84.55
Male e65	15.71	16.45	17.15	17.85	18.55
Female e65	20.06	19.46	19.96	20.46	20.96

Table 5. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Scenario 2	**2015***	**2025**	**2030**	**2035**	**2040**
Male e0	73.60	77.18	78.58	79.98	81.38
Female e0	81.60	82.87	83.42	83.97	84.52
Male e65	15.71	17.55	18.35	19.15	19.95
Female e65	20.06	23.01	23.76	24.51	25.26

Table 6. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Scenario 3	2015*	2025	2030	2035	2040
Male e0	73.60	78.37	79.87	81.37	82.87
Female e0	81.60	85.08	85.63	86.18	86.73
Male e65	15.71	19.55	20.25	20.95	21.65
Female e65	20.06	22.31	22.61	22.91	23.21

Table 7. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Projected life expectancy should not be discussed apart from official projections produced by Central Statistical Office of Poland (CSO, 2014) for 2015-2050. Mortality and life expectancy official projections are based on the target value (derived by comparison of life expectancy to selected countries from Western Europe), and were then calculated in three different scenarios of future development:

- medium scenario - "delay” of Polish mortality in relation to the developed countries will be maintained at the same level throughout the forecast period,
- low scenario - "delay" of Polish mortality will be remained at the same level until 2025; however, in subsequent years the reduction in mortality would be observed,
- high scenario - Polish mortality distance to the developed countries will gradually decline throughout the forecast period.

Under assumption that the future trend of life expectancy will be similar to trend observed for countries from cluster 1 (slow increase), we can expect that life expectancy at birth for female will increase from 81.60 in 2015 to 84.55 in 2040 (for male: from 73.60 to 80). According to the second scenario, where we assume that increase will be continued at a steady rate, life expectancy at birth for female will increase from 81.60 in 2015 to 84.52 in 2040 (for male: from 73.60 to 81.38). Acceleration in life expectancy (third scenario) will result in increase life expectancy at birth for female from 81.60 in 2015 to 86.73 in 2040 (for male: from 73.60 to 82,87). Meanwhile, according to the CSO of Poland, projected life expectancy at birth in 2040 increase to: (a) 85.9 (female) and 79.5 (male) under low scenario, (b) 86.5 (female) and 80.3 (male) under medium scenario, (c) 86.7 (female) and 80.9 (male) under high scenario.

Conclusions

We have applied a model that takes advantage of the past linear trends in life expectancy to make predictions about future life expectancy values. The model is based on block maxima method and uses generalized extreme distribution for modeling “best life expectancy” values.

Main advantage of this method is that it uses life expectancy levels for closely related populations. The method is worth further studying due to the reasonable results and – what is particularly interesting – comparable to the projections of NSO of Poland.

References

Assunção, R. M., Neves, M. C., Camara, G., & Freitas, C. D. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *Journal of Geographical Information Science, 20*, 797-811.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. *Springer Series in Statistics*.

Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika, 89*(2), 484-489.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling Extremal Events.

Fisher, R. A., & Tippett, L. H. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society, 24*(02), 180.

Gilli, M., & Këllezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics, 27*(2-3), 207-228.

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics, 44*(3), 423.

HMD. Retrieved February 04, 2017, from http://www.mortality.org/.

Lazar, D., Buiga, A., & Deaconu, A. (2016). Common stochastic trends in European mortality levels: testing and consequences for modeling longevity risk in insurance. *Romanian Journal for Economic Forecasting, 2*, 152-168.

Li, N., & Lee, R. D. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography, 42*(3), 575-594.

Medford, A. (2015). An Extremely different way to model and project Life Expectancy. Retrieved from http://epc2016.princeton.edu/uploads/160906

Oeppen, J., & Vaupel, J. (2002). DEMOGRAPHY: Enhanced: Broken Limits to Life Expectancy. *Science, 296*(5570), 1029-1031.

Population projection 2014-2050. (2014). Warsaw: CSO, Demographic Surveys and Labour Market Department. http://www.stat.gov.pl/

The 2015 Ageing Report. Underlying Assumptions and Projection Methodologies. (2014). *European Economy, 8*.