WAR EFFECT TO COHORT LIFE TABLES IN THE CZECH REPUBLIC

PETR MAZOUCH
University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Economic Statistics,
W. Churchill Sq. 4, Prague, Czech Republic
e-mail: mazouchp@vse.cz

Abstract

Nowadays cohort life tables have become more important for many reasons. Information about mortality patterns in the cohort perspective is relevant for insurance companies, pension funds, but also for the professional public in general. For a cohort life table, data of about at least 100 years are needed. Official data of mortality in the Czech Republic have periods with no or very poor information. The aim of this paper is to describe a possibility how to bridge one of those periods – the Second World War where there is a lack of information – and to discuss influence of short term periods on the mortality level of all cohorts. Real data are compared with models and influence to total life expectancy is measured.

Key words: *mortality patterns, cohort, life expectancy, the Second World War.*

1. Introduction

Construction of life tables is known for long time. Aim of the life table is to describe mortality patterns of some hypothetical population through some aggregated indicators. In 17th century J. Graunt found that there is different mortality rate in different age groups (Pavlík et al., 1986) and the first life table was constructed at the end of 17th century by E. Halley (Pavlík et al., 1986). Mortality patterns and studying and their development became very important for the whole society. Insurance companies use mortality tables to evaluate their products and governments use it to plan and apply some kinds of pension system.

Searching for the best model which could be used to describe mortality patterns in general is here from the beginning. We can find very old models as from Gompertz (1825) to some new techniques as presented by Mazouch (2009).

Application of mortality models can be found in areas as insurance industry. Zimmermann et al. (2014) used different mortality model for education groups to estimate difference in present value of retirement pensions for education groups. Fiala and Langhamrová (2014) applied mortality forecast to evaluate ability and stability of the social system in the Czech Republic from the view of revenues and expenditures. But mortality is not connected with population of man only, but we can find many useful applications of mortality models in other disciplines. Also Krejčí (2013) describes models of aging of machinery and equipment in education as aspect of modernity. The model was developed by Krejčí and Mazouch (2015) and applied in agriculture in estimation of age of machinery and equipment in the Czech agriculture (Krejčí et al., 2015).

Those models are based on information about some population and its development (with accent to mortality). Data needed for those models are usually the latest we have and we combine information about quantity of the population (divided by some characteristics as age, sex, education etc.) and quantity of events (from our point of view deaths – and also divided by the same characteristics). If data are from one year for the whole population we call this transversal model (Pavlík et al., 1986). This is the most common model to construct life tables. Those tables are constructed for one year and we use information about number of population from that year and number of deaths from the same year. Transversal life table combine information “what happened” in that current year to all population – through all ages, through all cohorts.

If the model is based on a cohort approach we do not use information about one year only but from many subsequent years through which cohort survived. Usually cohort is defined by one (or more) years of birth (Pavlík et al., 1986) and is very demanding in the data area.

To construct the whole life tables (from birth to the highest ages) we need data about complete cohort – extinct cohort. Time needed to die out one cohort is more than one hundred years and this problem causes that cohort life tables are available in some countries only. But cohort models are very important as one of the other source of information of mortality patterns (Mazouch, 2012).

In the Czech Republic complete cohort life tables absent. There are some fragments of the cohort analysis as for example Hulíková Tesárková and Mazouch (2013) who described basic cohort mortality analysis at higher ages (over 60) on cohorts born in 1890 – 1910 in the Czech Republic and compared results with France but the whole analysis is not still finished.

The problem is data. In the history of the Czech Republic there are some periods with no or very poor information about population or about deaths. If we watch history of the Czech Republic reversely the first problem is with The Second World War (WW II) when data are very poor (for more detail see Data part below). Aim of the paper is to describe data of the period of the WW II in the Czech Republic and its effect on cohort mortality indicators. Construction of the transversal life tables in this period is almost impossible (or very weak) but when we construct cohort life tables this period is only six (or seven if we include 1938 year to the analysis) years from more than one hundred. Analysis also shows influence of age when the cohorts are affected by WW II and describes basic model to bridge this period in the cohorts’ life times.

2. Data

Before description of data source we have to explain possibilities of data classification. The Czech demographic statistics is of a good quality and it was also in the past. But for different time periods the data for cohorts have to be taken from different sources which are often classified in a different manner.

Lexis diagram (for more about Lexis diagram see Pavlík et al., 1986) easily illustrates our data. Each person in population starts his or her line on axis *x* (year) at age $x = 0$ (axis *y*) and follows trend under 45 grades to right. At the moment of death the person is in some area (defined as triangle) at some exact age (on axis *y*) and some exact time (on axis *x*). Cohort is defined as group of persons born in one year (their life lines start for example between points B and C).

Number of deaths can be defined in three different ways:

- 1. set of events defined as all deaths in area within points A, D, F and E,
- 2. set of events defined as all deaths in area within points B, D, E and A,
- 3. set of events defined as all deaths in area within points B, C, D and A.

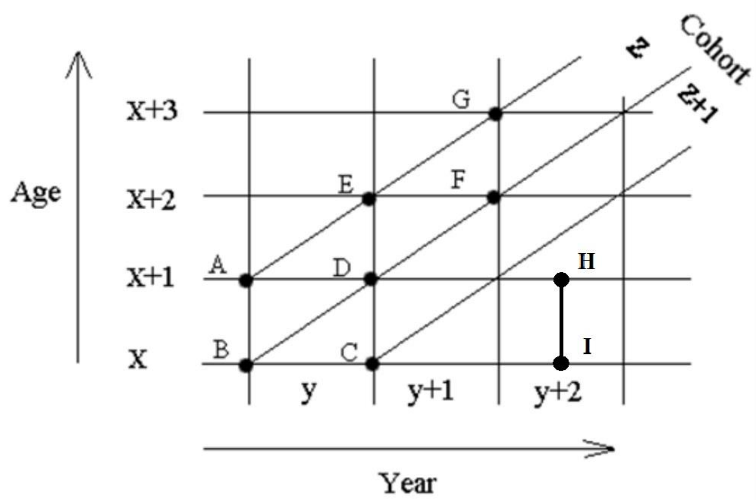
The first set of events covers deaths defined by one cohort, one year of age but covers two calendar years. The second set of events is defined by one cohort, one calendar year when events occurred and two age groups. The third set (square) of events covers deaths defined by one year of age, one calendar year but two cohorts. All three sets are based on triangles and if we know the number of events in triangles we can then join two neighbouring triangles to any set of events (for example triangles BCD and BDA give 3. set of events, BDA and ADE give second, etc.).

To calculate probability of dying or mortality rate number of exposed population is needed. Number of exposed population is in Lexis's diagram expressed as group of intersections of life lines and some vertical line.

We have two ways how to estimate exposed population:

- number of survivors to the end of the year defined as line between points C and D,
- number of survivors to the middle of the year defined as line between points I and H.

Figure 1: Lexis diagram



Source: adapted from Pavlik et al. (1986).

Data in this paper covers period between year 1920 and 2008. For the whole period number of deaths is classified in basic triangles what means that we can construct any set of events. For different time period data do not correspond to the same territory (will be discussed later). For exposed population data are from 1945 till 2008 as number of survivors to the end of the year and before the year 1945 (between 1920 and 1945) data are to the middle of the year.

The problem with different indicators in different periods is not the most important. In period from 1920 to 1938 and from 1945 till the present region of the Czech Republic was the same and data from this period (number of deaths, population) are comparable.

Different situation is in period from 1938 to 1945. At the end of 1938 German empire started to occupy Czech regions next to boarders. Those regions became part of German empire and the rest – central part of Czech Republic became Protectorate of Bohemia and Moravia (Kučera, 1994).

Statistics from this period are incomplete and the quality of data is worse. Estimates of migration suppose that around half from almost one million of Czech citizens in area near to boarders moved to Protectorate. German population in Protectorate was around 3 % (Kučera, 1994). There are no data about population in this period. All estimates are done from year 1945 on the base of balance of population without migration (population from the beginning of 1945 plus deaths in 1944 minus new births in 1944 gives population at the beginning of

1944). More details are discussed in Kučera (1994). From this method we have estimated population structure for Protectorate and for Czechs only.

Statistics of events – deaths is also incomplete. Kučera (1994) mentions that mortality during WW II is slightly increasing but not rapidly (at the end of War only when some battles were in the Czech Republic). In official statistics only records for Czech citizens in Protectorate are available (Population change 1938 – 1945). Around 77 thousands of Jewish and 55 thousands of other citizens who died in concentration camps and in prison or were killed during War are included in official statistics only partly (Kučera, 1994).

Lack of information about events and about population could cause under estimates of real mortality rate during WW II. For this reason real (incomplete) data are compared with modeled data. It is clear that there is difficult situation to construct transversal life tables to know mortality patterns in exact year during WW II but cohort life tables construction is different and this imperfection of data is not so essential because it is period of 7 years only in the long period of life of the cohort. Effect of the higher level of mortality to cohort indicators is much less then to transversal indicators.

For our analysis data available for generation from 1860 to 1930 are used. The oldest generation is affected by War at age around 80 years (from 79 to 85) and the youngest at very young ages (from 9 to 15 years). As data are from 1920 only the life tables are shorted and start at age when population was at year 1920 (for younger generation born later than 1920 tables start at age 5 to eliminate early child mortality) and for population which are not extinct the limit is year 2008 (for not extinct cohorts the last age is when the cohort reached year 2008).

3. Methodology

As the first step in methodology data about deaths are structured to the first and second sets of events (see Section 2). This is possible for data structured to the triangles and as was mentioned in Section 2 this data are available for the whole period (from 1920 to 2008 – see the Czech Statistical Office, 2010).

Data of exposed population are available in two ways. The first is number of survivors at the end of the year (for period 1945 to 2008) and the second is the number of survivors to the middle of the year (1920 to 1945). For the analysis only data about number of survivors at the end of the year are useful. According to Chiang (1984) estimates of number of survivors at the end can be estimated as average from two neigbouring number of survivors to the middle of the year at age x.

Effect of the Second World War will be illustrated by probability of dying counted from the second sets of events divided by population at the end of the year. Mortality rates are counted from the first sets of events divided by population at the end of the year. Mortality rates are then used for mortality tables for each generation. Methodology of life tables construction used in the paper respects Chiang (1984).

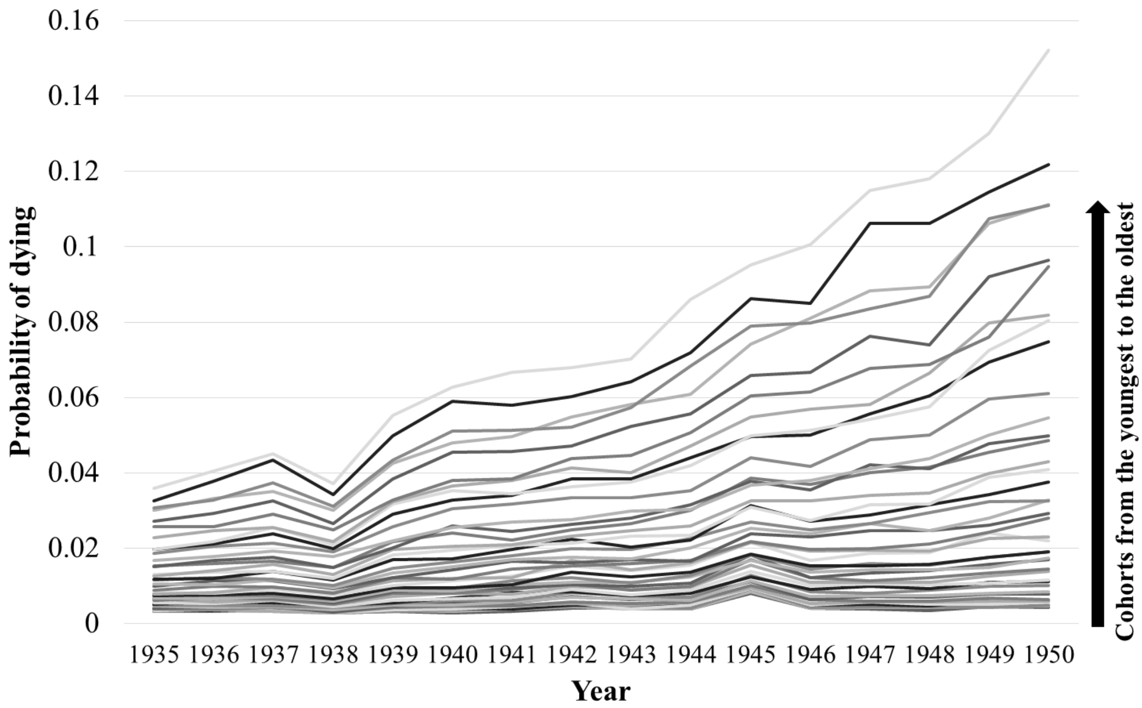
Because of lower credibility of data from the period of WW II some models of smoothing are used. For each cohort period war was modeled by very simple assumption of constant growth rate of the mortality rate. The relative difference of the last mortality rate before 1938 (it differs for cohort – for example for cohort 1860 it is age 76, for cohort 1861 age 77 etc.) and the first mortality rate after 1945 (for cohort 1860 it is age 86, for cohort 1861 age 85 etc.) gives total growth of the mortality between those ages. The tenth root extraction of this total growth gives us annual average growth of mortality rate between those two ages.

Real results (with original incomplete data) are then compared with smoothed model data. To estimate the impact of the war smoothed mortality rates are increased by constant (only in the period of the war) and influence of this increase is measured. As the basic indicator Life expectancy (Pavlík et al., 1986) is used.

4. Results

As Fig. 2 illustrates there are some remarkable changes in mortality during the Second World War. That is important to notice that variability could be caused not only by variability of data itself but also by data quality as this was discussed in Data part. Fig. 2 covers only part of studied cohorts and also part of studied time period. The cohort restriction is because wider cohort range we include to the graphical analysis less detailed information we get. Only cohorts between 1890 and 1930 are covered (cohort 1890 was during the War at age around 50, younger cohorts less).

Figure 2: Probability of dying between 1935 and 1950 for cohorts 1890 – 1930, Czech Republic, Males, second sets of events



Source: the author based on data from the Czech Statistical Office (2010).

The first change in the trend is decline in 1938 what could be caused by data quality. It is remarkable for all cohorts. The second move in opposite direction is in year 1945. This could be caused by both – higher mortality and data quality. Higher mortality during the end of the war was mentioned above already. Small increase of the mortality can be visible also in 1942 but the increase is much lower than in 1945.

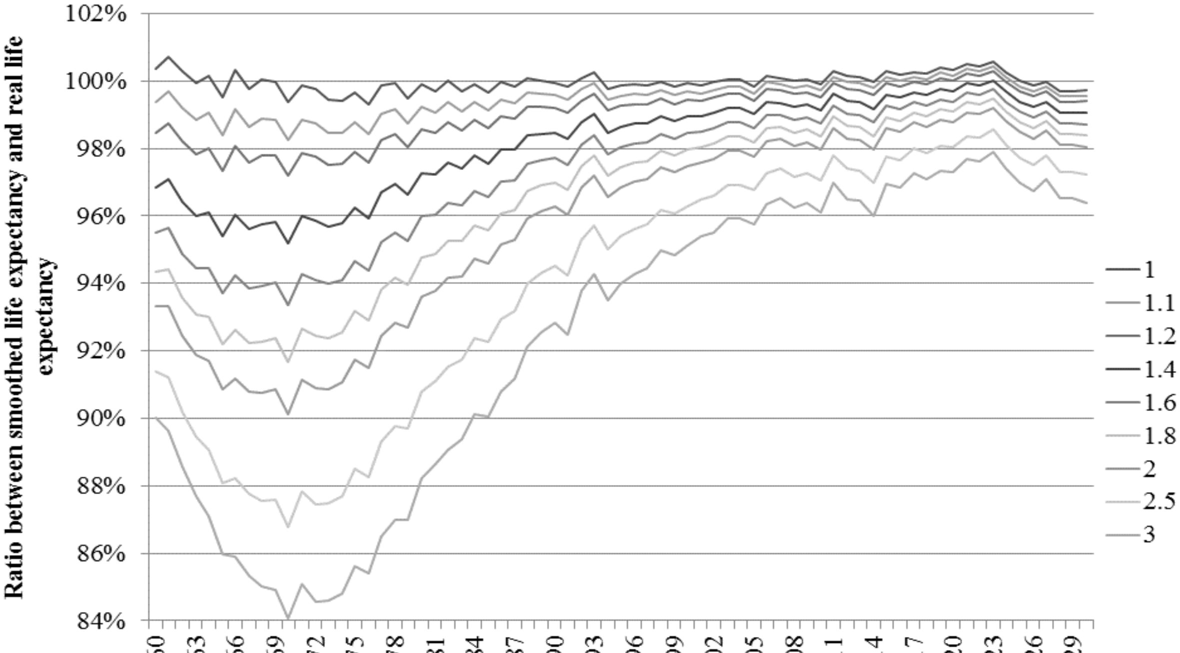
Increase in 1945 is for some cohort on the level of one third, the question is, how this increase could influence total life expectancy of the whole cohort. Analysis of older cohorts show that the impact of the year 1945 is declining and for cohorts older than 1880 the effect is not there anymore (on the other hand decrease of the year 1938 is there still). The same results are for women but the increase in 1945 is lower for majority of cohorts than for man.

It is also visible that for no cohort there is any high increase of the mortality during the WW II for all the period. Trends between year 1937 and 1946 are higher slightly only and the effect of this increase is the question.

To evaluate effect of the war smoothed data were estimated (for more details see Methodology part). If we estimate no change in the mortality development during the war period the mortality rate between two ages would increase constantly (in relative expression). Difference between the real data and this smoothed mortality rates shows us the effect of the war. Results are in Fig. 3 and 4 and the relative differences are described by the highest line (coefficient 1 – constant relative increase during the war). The line is only very lightly different from the 100 % line. It could be explained as that the real data do not express any high impact of the war to mortality in cohort perspective.

In data section problem with real data was mentioned already. We know that data from the war are probably underestimated and for sure very variable. The model in the Fig. 3 and 4 shows impact of hypothetical increase of the mortality level during the war. Illustrated variants are increase of 10 % (coefficient 1,1), 20 % (coefficient 1,2), 40 %, 60 %, 80 %, 100 % (coefficient 2), 150 % and 200 % (coefficient 3). Coefficient 1 is for smoothed data with no theoretical increase.

Figure 3: Ratio between smoothed cohort life expectancy and real cohort life expectancy for cohorts from 1860 to 1930, variants of increase of 10 %, 20 % (coefficient 1,2), 40 %, 60 %, 80 %, 100 % (coefficient 2), 150 % and 200 % (coefficient 3), Czech Rep., Females



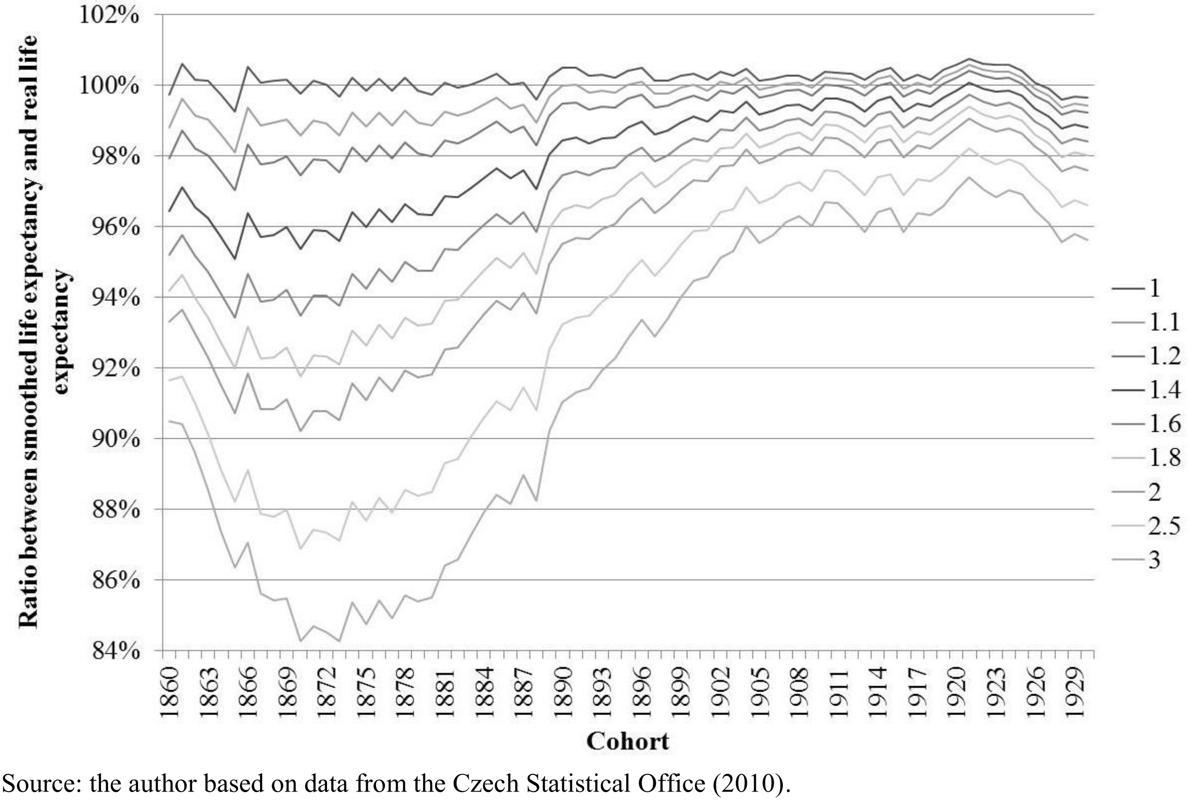
Source: the author based on data from the Czech Statistical Office (2010).

If the effect of the war would be for example 40% increase of mortality level (for all years from 1938 to 1945) from the theoretical model "without war impact" we can see for different cohorts the different impact. The highest is for females cohorts born in late 60. in 19th century up to cohorts 1890. This is clear because those cohorts were during the period of war at age with the very high mortality rate and age with the highest number of deaths. We can see that higher the coefficient is the higher the difference is. But for cohorts born after 1900 the difference in life expectancy is very small (up to 2 % if we talk about the 40% increase of mortality level during the war).

The interpretation of the results is very difficult because we do not have any complete cohort in our dataset (we start with data from 1920 to 2008) but for example cohort born in 1910 (we have life expectancy at 10 years for this cohort) was between 28 and 35 during the WW II. This cohort is nowadays almost extinct and the cohort life expectancy (at age 10) for this cohort from real data is 62.15. If there is no war and mortality in this period is increasing constantly the life expectancy would be 62,2 and if we estimate increase of the mortality level of 40 % from the "ideal" the life expectancy decline to 61,7. The difference is less than 0.5 % of the life expectancy.

For males results are very similar with only one difference. The group of cohorts where the effect of higher mortality is visible is wider. This is caused by male's different mortality development when the age with higher mortality is wider than for females.

Figure 4: Ratio between smoothed cohort life expectancy and real cohort life expectancy for cohorts from 1860 to 1930, variants of increase of 10 %, 20 % (coefficient 1,2), 40 %, 60 %, 80 %, 100 % (coefficient 2), 150 % and 200 % (coefficient 3), Czech Rep., Males



Source: the author based on data from the Czech Statistical Office (2010).

5. Conclusion

It is known that poor data quality could cause questionable results. Data from the Second World War gives us incomplete information about mortality patterns during the war. This is the reason why construction of the life tables from this period needs some estimates and models and for transversal type of life tables the data are not convenient at all.

The period of the war (including years before the war also) when the data are very poor was 8 years. For cohort this is "only" eight years from the total life span (usually for cohorts more than 100 years). Possibility how to estimate mortality patterns during this period could be done by simple model of smoothed mortality rates with assumption of no impact of the war. Comparison shows that real data (with all its imperfection mentioned in the paper) do not show any big war impact.

On the other hand it is known that some impact of the war increase the mortality level. Estimates of some non-recorded events indicates that increase of the mortality could be in tens of percent but not much higher (the highest peak in mortality was in 1945 and for some cohorts the increase was about 50 %). This could cause decline of the life expectancy in some specific cohorts (which were during the war at age with high number of deaths – years between 50 for males and 60 for females and 80). Younger cohorts were affected by war in cohort perspective only slightly.

Next research would be focused on prolongation data set to the 1. World War where similar situation to period of the 2. World War is expected. Longer data sets could give us more information about complete generations and could quantify the war effect better.

Acknowledgements

The paper has been prepared under the support by Grant Agency of the Czech Republic to the project no. P404/12/0883 "Generační úmrtnostní tabulky České republiky: data, biometrické funkce a trendy".

References

[1] CHIANG, C. L. 1984. The life tables and its applications. Malabar : Krieger, 1984.

[2] CZECH STATISTICAL OFFICE 2010. Population change 1920-2008. Prague : Czech Statistical Office, 2010.

[3] FIALA, T., LANGHAMROVÁ, J. 2014. Modelování budoucího vývoje úhrnu pojistného a úhrnu vyplacených starobních důchodů v ČR. In Politická ekonomie, 2014, vol. 62, iss. 2, pp. 232-248.

[4] 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. In Philosophical Transactions of the Royal Society of London, 1825,vol. 115, pp. 513-585.

[5] HULÍKOVÁ TESÁRKOVÁ, K., MAZOUCH, P. 2013. Basic cohort mortality analysis at higher ages: an analysis of the rectangularisation process based on cohorts born in 1890 – 1910 in the Czech Republic and France. In Demografie, 2013, vol. 55, iss. 1, pp. 27-46.

[6] KREJČÍ, I. 2013. Age of machinery and equipment in education as aspect of modernity. In Proceedings of the 10th International Conference on Efficiency and Responsibility in Education 2013 (ERIE 2013) 06.06.2013, Prague, Czech Republic. Prague : Czech University of Life Science Prague, 2013. pp. 317-323.

[7] KREJČÍ, I, MAZOUCH, P. 2015. Age of equipment in education – a possible indicator of the modernisation process. In Proceedings of the 12th International Conference Efficiency and Responsibility in Education 2015 (ERiE), Praha, 04.06.2015 – 05.06.2015. Prague : Czech University of Life Sciences Prague, 2015, pp. 289-296.

[8] KREJČÍ, I. et al. 2015. Age of machinery and equipment in the Czech agriculture. In Agricultural Economics, 2015, vol. 61, iss. 8, pp. 356-366.

[9] KUČERA, M. 1994. Populace České republiky 1918 – 1991. In Acta Demographica, 1994, vol. XII, pp. 198.

[10] MAZOUCH, P. 2009. Modelling of mortality in Czech Republic. In AMSE 2009 Applications of Mathematics and Statistics in Economy. Uherské Hradiště, 26.08.2009 – 28.08.2009. Praha : Oeconomica, 2009, pp. 289-296.

[11] MAZOUCH, P. 2012. Possibilities of cohort mortality modelling. In AMSE 2012 Applications of Mathematics and Statistics in Economy. Liberec, 30.08.2012 – 01.09.2012. Praha : University of Economics, Prague, 2012, pp. 9.

[12] PAVLÍK, Z., RYCHTAŘKOVÁ, J., ŠUBRTOVÁ, A. 1986. Základy demografie. Praha : Academia, 1986.

[13] ZIMMERMANN, P., MAZOUCH, P., TESÁRKOVÁ HULÍKOVÁ, K. 2014. The difference in present value of retirement pensions for education groups. In The 8th International Days of Statistics and Economics. Prague, 11.09.2014 – 13.09.2014. Slaný : Melandrium, 2014, pp. 1715-1721.