# Identifying the Ruptures Shaping the Segmented Line of the Secular Trends in Maximum Life Expectancies

Carlo G. Camarda[1], Jacques Vallin[2], and France Meslé[2]

[1]Max Planck Institute for Demographic Research
[2]Institut national d'études démographiques

Extended abstract submitted to the EPC 2012
October 2011

## 1 Background

In an impressive article published in Science in 2002, James Oeppen and James Vaupel showed that the secular trend in the highest life expectancy reached each year in the world follows an straight line (R=0.992) with a 25% gradient, indicating that the maximum life expectancy observed had increased steadily by three months per year, since 1841 (the start year of their database). If we had been nearing a limit, they wrote, we would have seen a flattening of the curve over recent decades. As this is not the case, there is no reason to believe that progress in life expectancy will stall in the near future.

More recently, Vallin and Meslé (2009, 2010), by looking at more data (more life tables for the 1841-2000 period and additional tables before 1841 and after 2000) but, more importantly, by checking better data quality, found that maximum life expectancies do not follow a straight line but a succession of four segments, corresponding to four important changes in the history of the health transition in the most advance countries (Figure 1).

The first cut occurred in the years 1790s, at the time of Jenner and of the French Revolution, the second around the year 1885 at the time of Pasteur and the third at the end of the 1960s when started the so-called Cardiovascular Revolution. Not only correlation coefficients of each segment are higher than those of Oeppen and Vaupel, but their slopes differ (0.5%, 11%, 32%, and finally 23%), indicating that at each stage correspond new tools for health improvements that results in new paces of life expectancy increase. After plateauing at very low levels for thousands of years, maximum life expectancy started increasing at the eve of the 19th century, at a modest but steady pace thank to important progresses in food availability and first modern tools to fight against infectious diseases. Pace of increase suddenly accelerated in the mid 1880s with the Pasteur revolution. In the

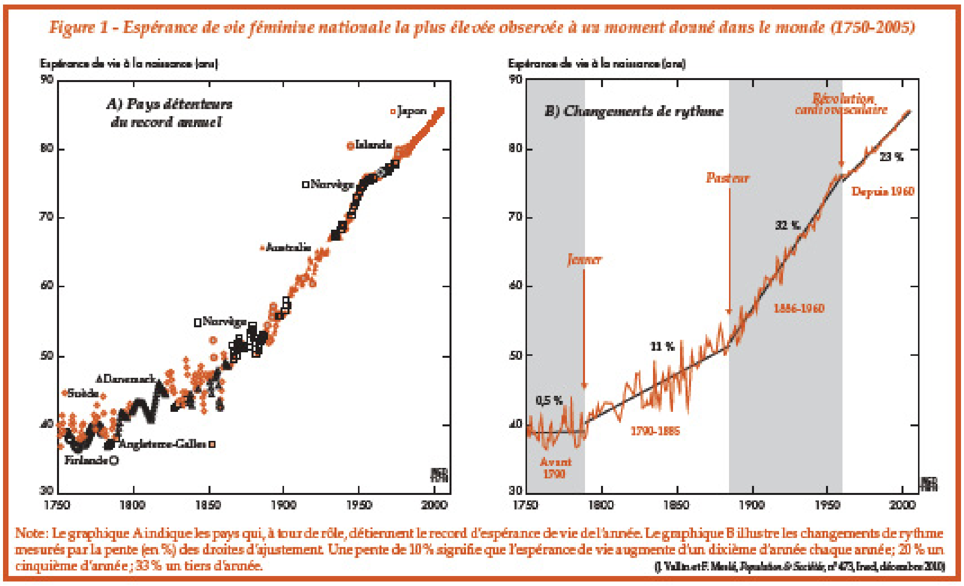


Figure 1: Maximum life expectancies over time (1750-2005). Left panel: countries represented in the dataset are depicted by different point-types. Right panel: regression lines with associated breaks. Source: Vallin and Meslé (2010).

1960s however, the benefit of receding infectious diseases came to their end in the most advanced countries and if life expectancy finally do not stop to increase is mainly due to the discovery of new ways to fight cardiovascular diseases it nowadays increases at a reduced pace since life expectancy gains more and more rely on reducing mortality at older and older ages. Furthermore, to their turn, the benefits of cardiovascular diseases receding is no longer far to their end in the most advanced countries and the near future depends again on the emergence of new innovation of that we do not know but that will probably result in a new change in pace of life expectancy increase, if any.

However, Vallin and Meslé's identification of the time cuts was intuitive, simply suggested by the general shape of the cloud of points. The objective of this paper will be to check if it is possible to confirm both the number of cuts and their calendar by more objective statistical methods.

## 2 Segmented regression

Instead of searching for a single regression line over time, or for a series of regression lines independently estimated within subjective chosen time-window, we assume that relationship between the maximum life expectancies and our explanatory variable, years, is piecewise linear, namely represented by two or more straight lines connected at unknown values. Segmented or broken-line regression models are an elegant framework for this type of setting and they allow the simultaneous estimation of the eventual breakpoints as well as the regression coefficients associated to each piece.

Briefly, assuming a single change-point, a segmented relationship between the response, $e_i^m$, maximum life expectancies, and the variable $y_i$, years, with $i = 1, 2, \dots, n$ is modelled by the following term:

$$e_i^m = \alpha + \beta_1 \, y_i + \beta_2 \, (y_i - \psi)_+$$

where $(y_i - \psi)_+ = (y_i - \psi) \cdot I(y_i > \psi)$ and $I(\cdot)$ is the indicator function equal to one when the statement is true. According to such parameterization, $\alpha$ is the common intercept, $\beta_1$ is the slope of the first segment, $\beta_2$ is the difference-in-slopes for the second segment and $\psi$ is the breakpoint. Figure 2 gives a schematic overview of a segmented regression model. Details of this class of models and the associated estimation procedure can be found in Muggeo (2003).

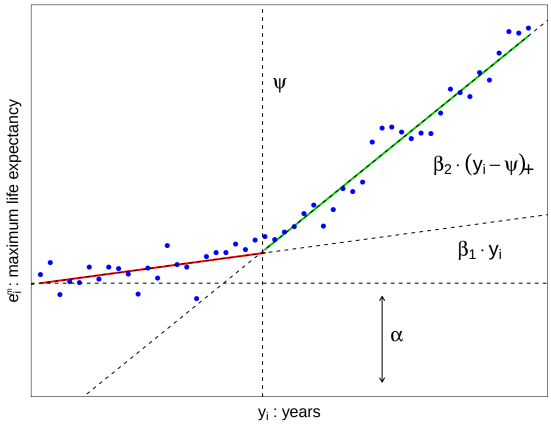


Figure 2: A schematic representation of a segmented regression model.

It is important to underline that we consider maximum life expectancies as a continuous variable over the whole time-window with breaks (i.e. a regression function which is continuous, but first derivatives discontinuous). Obviously record life expectancy experienced peculiar events, but there is not reason to see sudden jumps in the overall series, especially looking at the maximum achieved at any given year. Also with new tools for health improvements it is more natural to assume only a change in the slope. In this setting segmented regression allows the estimation of such slopes associated to each segment, and break-points with confidence intervals.

Finally linear model can be seen as a special case of a segmented regression model in which the change-point $\psi$ does not exist, i.e. $\beta_2$ is not significantly different than zero.

# 3 An application

Figure 3 presents a preliminary outcome of the segmented regression model applied to the data in Figure 1. The number of break-points was optimized by Bayesian Information Criterion (Muggeo, 2003) and found equal to three. Moreover the changing years recognized by Vallin and Meslé (2009) are within our estimated confidence intervals. We also found comparable gradients for each segment: 0.3%, 11.8%, 32.37% and 23.21%.

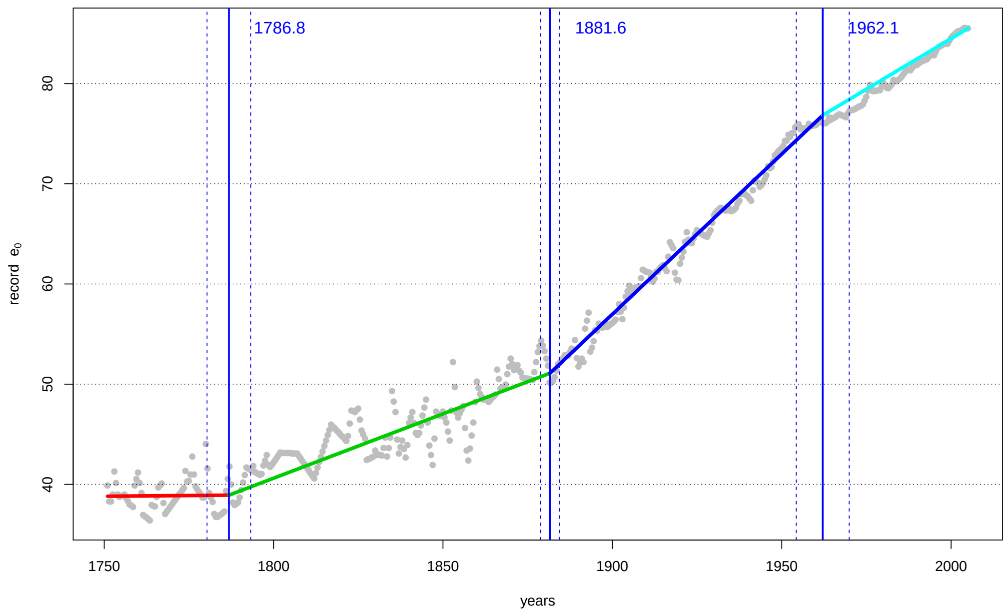


Figure 3: Actual and fitted maximum life expectancies over time by segmented regression model (1750:2005)

Given the importance of the research question, more work needs to be done for conforming the mentioned outcomes. Nevertheless we already see that segmented regression offers an elegant and suitable framework for statistically identify possible ruptures in the time series of maximum life expectancies. Further analysis by age will be also carried out to shed light on determinants of the eventual variations in gradient.

# References

Muggeo, V. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine 22*, 3055–3071.

Oeppen, J. and J. W. Vaupel (2002). Broken Limits to Life Expectancy. *Science 296*, 1029–1031.

Vallin, J. and F. Meslé (2009). The segmented trend line of highest life expectancies. *Population and Development Review 35*, 159–187.

Vallin, J. and F. Meslé (2010). Espérance de vie: peut-on gagner trois mois par an indéfiniment? *Population & Sociétés 473*, 1–4.