Li Qiu
VEE – Time Series
Student Project – Spring 2010
August 20, 2010

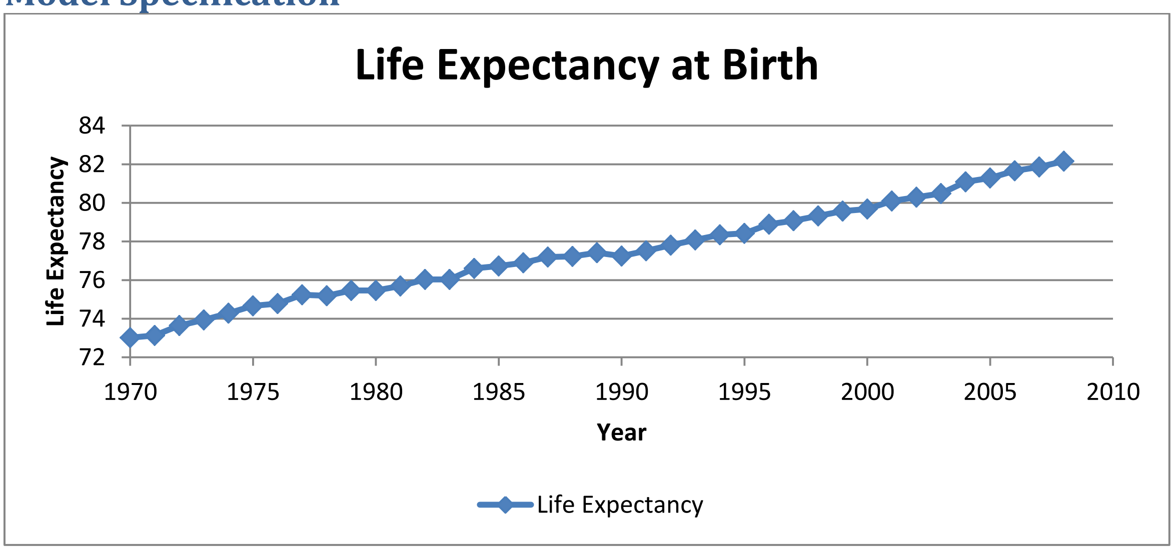
Introduction

The intention of the project is to model the life expectancy at birth of an individual in Switzerland. Several time series models are used to fit the first differences of the life expectancy at birth, including AR(1), MA(1) and ARMA(1,1). The goodness of fit will be estimated in the later sections of the project.

Data

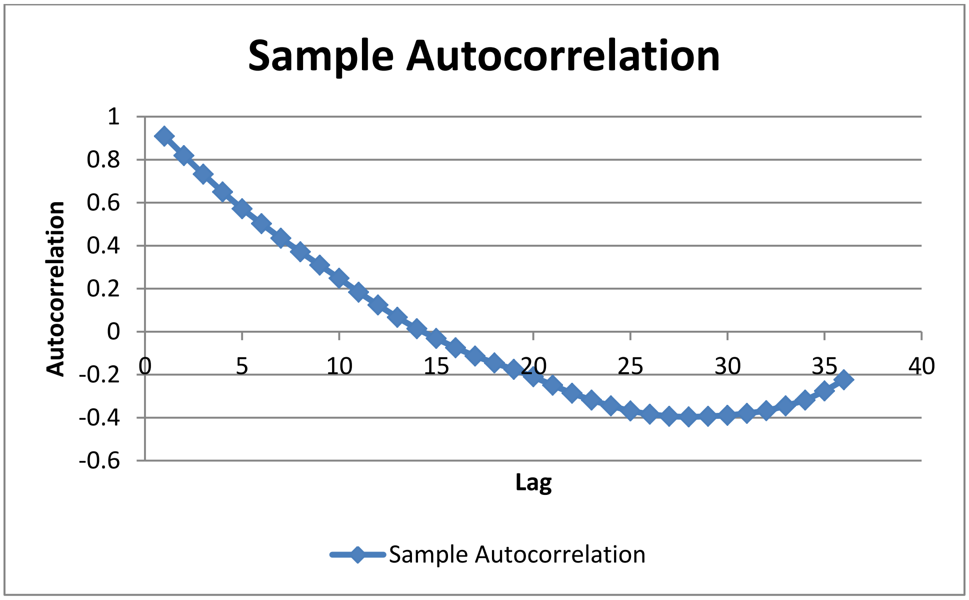
Data on the life expectancy in Switzerland were yearly observations from 1970 to 2008. The life expectancy at birth is the number of years a newborn infant can expect to live if the current patterns of mortality at the time of its birth were to remain constant throughout its life. The data is obtained from the World Bank, which in turn derives its sources from "(1) United Nations Population Division. 2009. World Population Prospects: The 2008 Revision. New York, United Nations, Department of Economic and Social Affairs (advanced Excel tables), (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) Secretariat of the Pacific Community: Statistics and Demography Programme, and (5) U.S. Census Bureau: International Database."
http://data.worldbank.org/indicator/SP.DYN.LE00.IN?cid=GPD_10

Model Specification



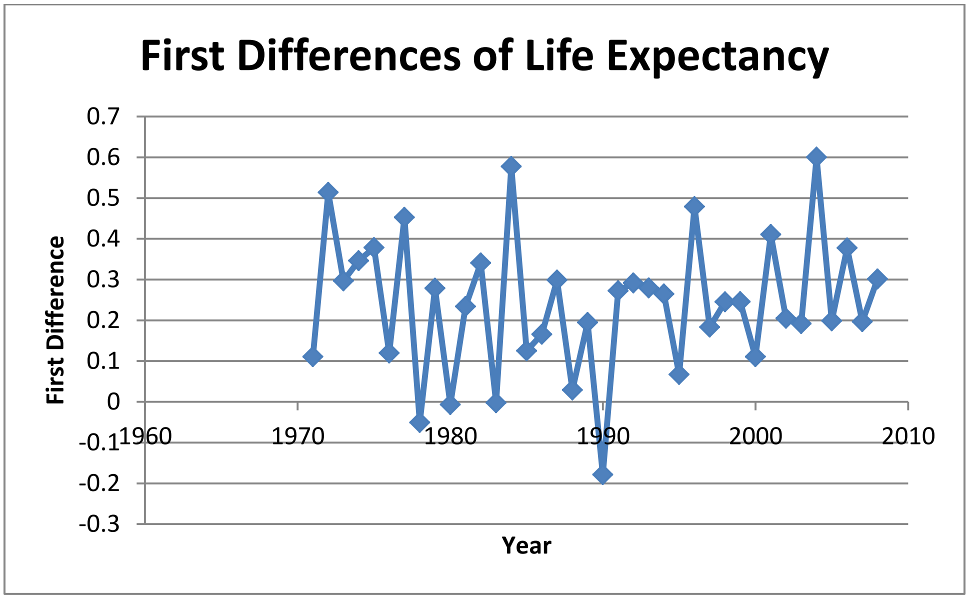
The life expectancy in Switzerland seems to generally trend upwards, with occasional dips or pauses observed in 1978 and 1990. Over the entire period, the average life expectancy increased by 9.14 years, or 12.5%, which equates to an 1.4% average annualized growth.

Being a directionally upwards trend, it does not appear to exhibit seasonality, and the growth rate appears to be fairly constant (not exponential).

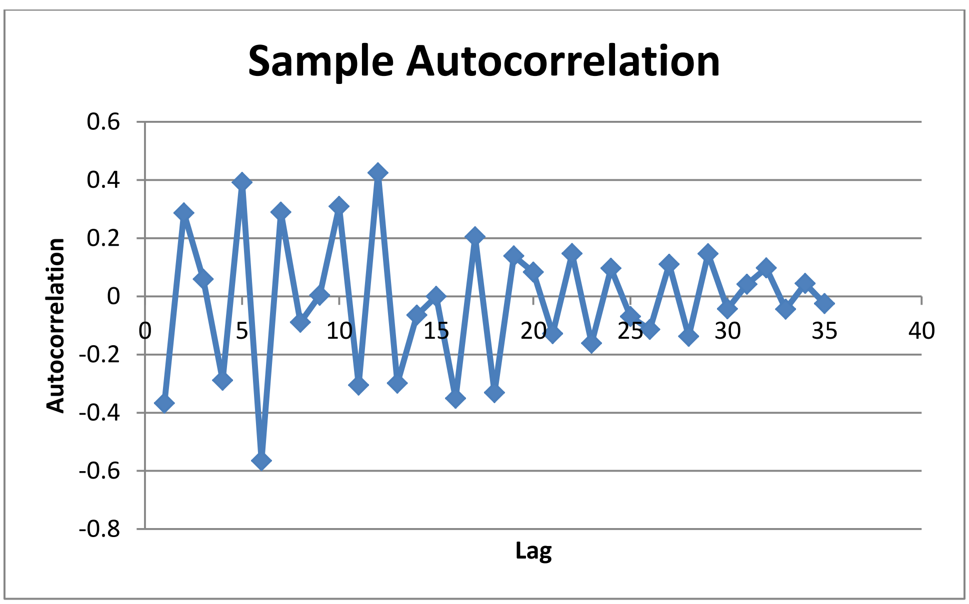


The autocorrelation values are calculated using the macro in the NEAS excel template technique spreadsheet. Looking at the sample autocorrelations in the correlogram, we see the autocorrelations initially decrease as lag increases to 28, then subsequently increase as lag continues to increase. Since a stationary series is expected to have autocorrelations that move rapidly to zero as lag increases, I conclude that the life expectancy in years is not a stationary series.

First Differences

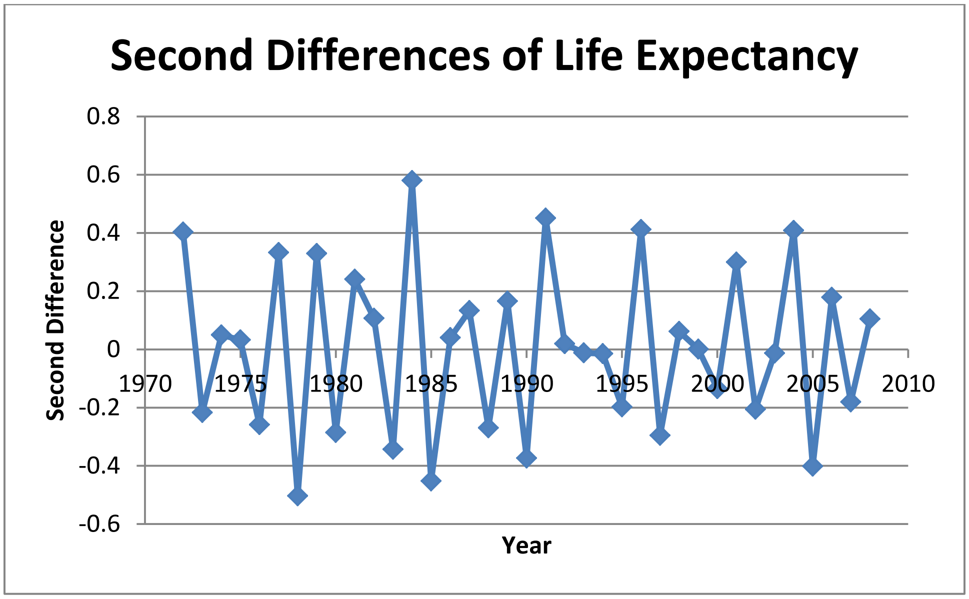


Taking first differences, I note that the first differences of life expectancy seem to oscillate about 0.25 years, which is fairly similar to the average growth of 0.24 years in life expectancy over the entire period. With the exception of outliers in 1984, 1990 and 2004, the entire series appears to be fairly stationary.

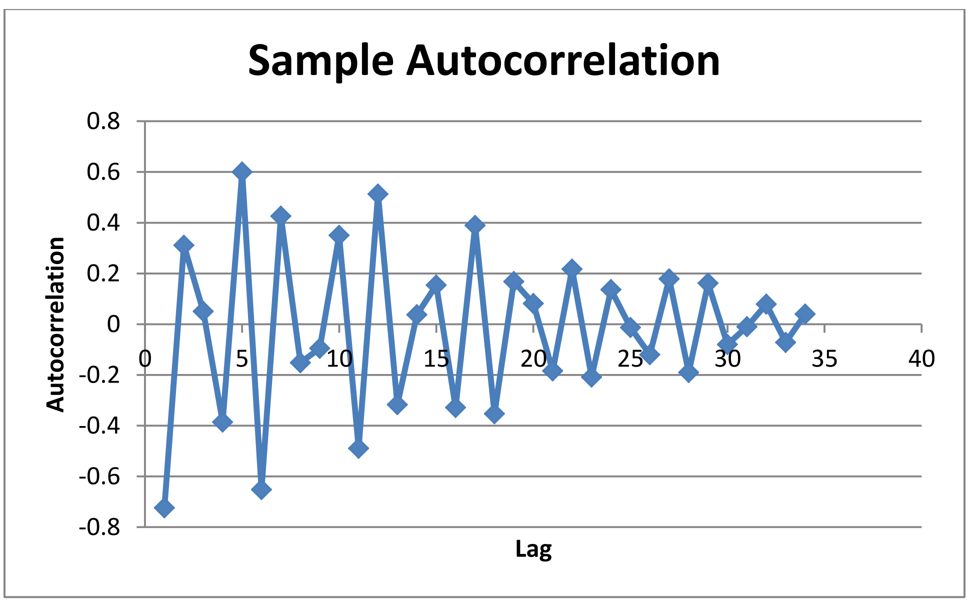


The correlogram for the sample autocorrelation in first differences show the sample autocorrelations oscillating about zero with decreasing amplitudes; this pattern suggests that the first differences are stationary. Since the autocorrelations persist at non-zero values for some time, I conclude that the process might be autoregressive in nature.

Second Differences



The second differences for life expectancy appear to be fairly similar in pattern with the first differences: oscillating around a zero mean. This series appears to be stationary as well.



The correlogram for the sample autocorrelation in second differences continues to show the sample autocorrelations oscillating about zero with decreasing amplitudes; this pattern suggests that the second differences are also stationary. However, the magnitude of the autocorrelations for the second differences seems to be higher when compared to those for the first differences. Since the autocorrelations persist at non-zero values for some time, I conclude that the process might be autoregressive in nature.

Parameter Estimation

AR(1) Model

Based on the regression output in the supporting Excel document, the AR(1) equation is

$$Y_t = -0.369Y_{t-1} + 0.332$$

The p-value is 0.0237, which means that the estimated ϕ is significant at the $\alpha = 0.05$ level. Unfortunately, the R-squared is only 0.138, which means that the model is only able to explain 13.8% of the observed variation in the first differences. Despite the low R-squared value, the autoregressive process is stationary as $|\phi| < 1$.

MA(1) Model

Using the method of moments and the Yule-Walker equation, I calculated

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1} = 0.438$$

Hence, the estimated MA(1) equation is:

$$Y_t = 0.241 + e_t - 0.438e_{t-1}$$

ARMA(1,1) Model

To obtain an estimate for ARMA(1,1), the method of moments was applied again.

By the ratios of the autocorrelations,

$$\hat{\phi} = \frac{r_2}{r_1} = -\frac{0.368}{0.287} = -0.780$$

Using the formula $\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-2\phi_1\theta_1+\theta_1^2}$, and backsolving,

$$\hat{\theta} = -0.496$$

Therefore, the estimated ARMA(1,1) equation is:

$$Y_t = -0.780Y_{t-1} + e_t + 0.496e_{t-1}$$

The ARMA model is also stationary, as it fits the stationary condition of $|\hat{\phi}| < 1$.

Model Diagnostic

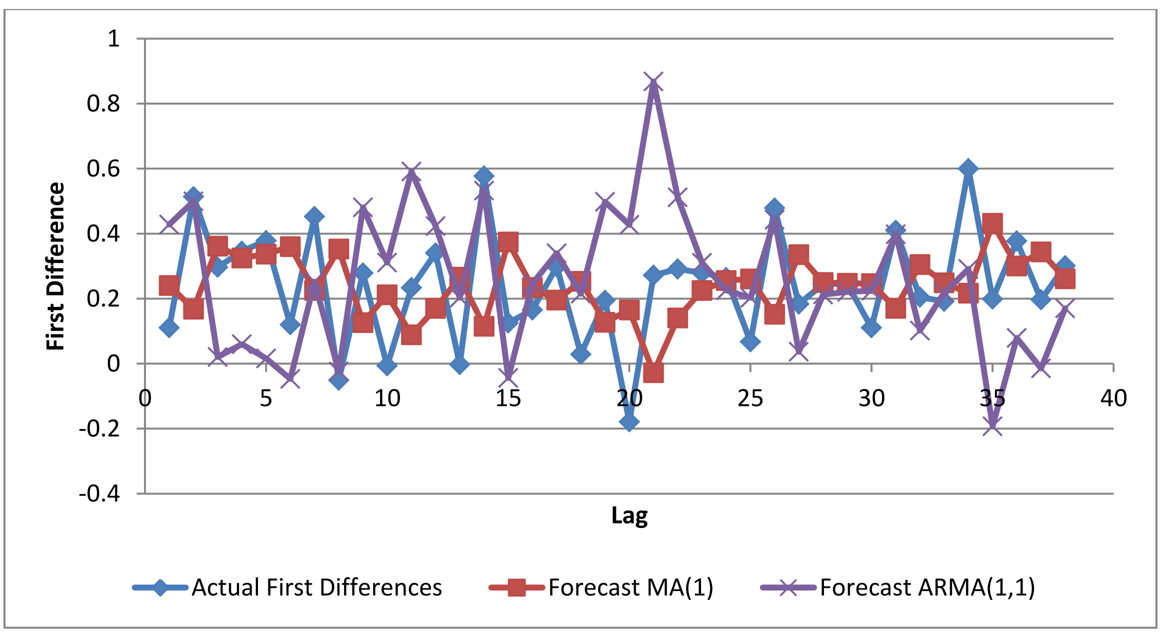
To determine the goodness of fit of these models, the Box-Pierce Q Statistic was calculated for 35 lags using the Time Series Techniques macro, and the results are shown in the table below.

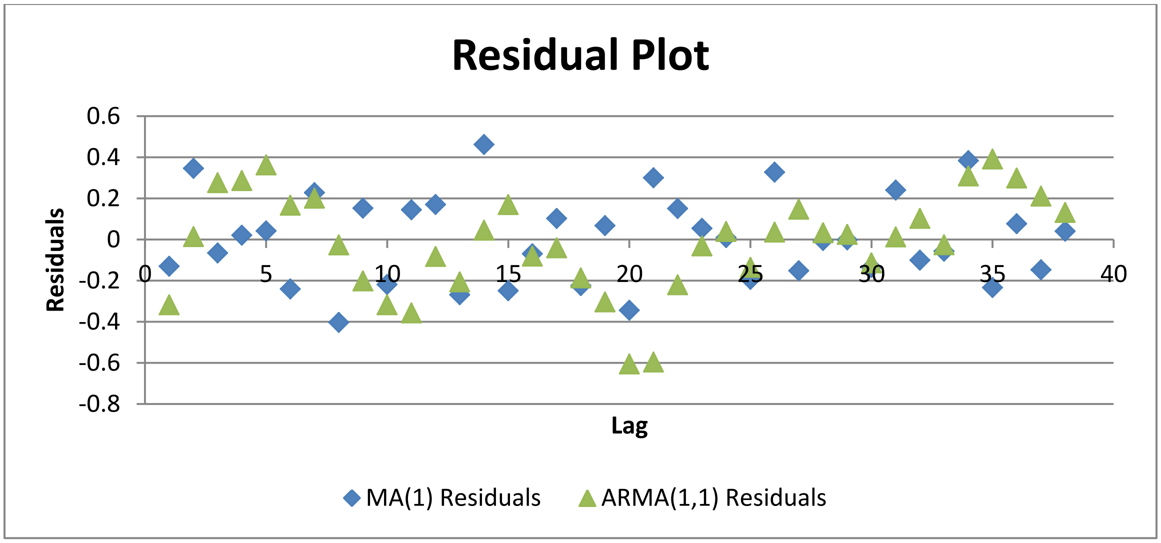
Model	Statistic
AR(1)	61.54635
MA(1)	42.57834
ARMA(1,1)	43.35124

The Box-Pierce test tries to assess if the autocorrelations of a time series are different from zero, and the null hypothesis would be that the residues are white noise. For lag 38, the critical value of the statistic is 44.903. Since the Q statistic is lower for both the MA(1) and ARMA(1,1) models, we can reject the null hypothesis at the $\alpha = 0.10$ level. However, we are unable to reject the null hypothesis for the AR(1) model, which suggests that it is probably not a good fit for the data.

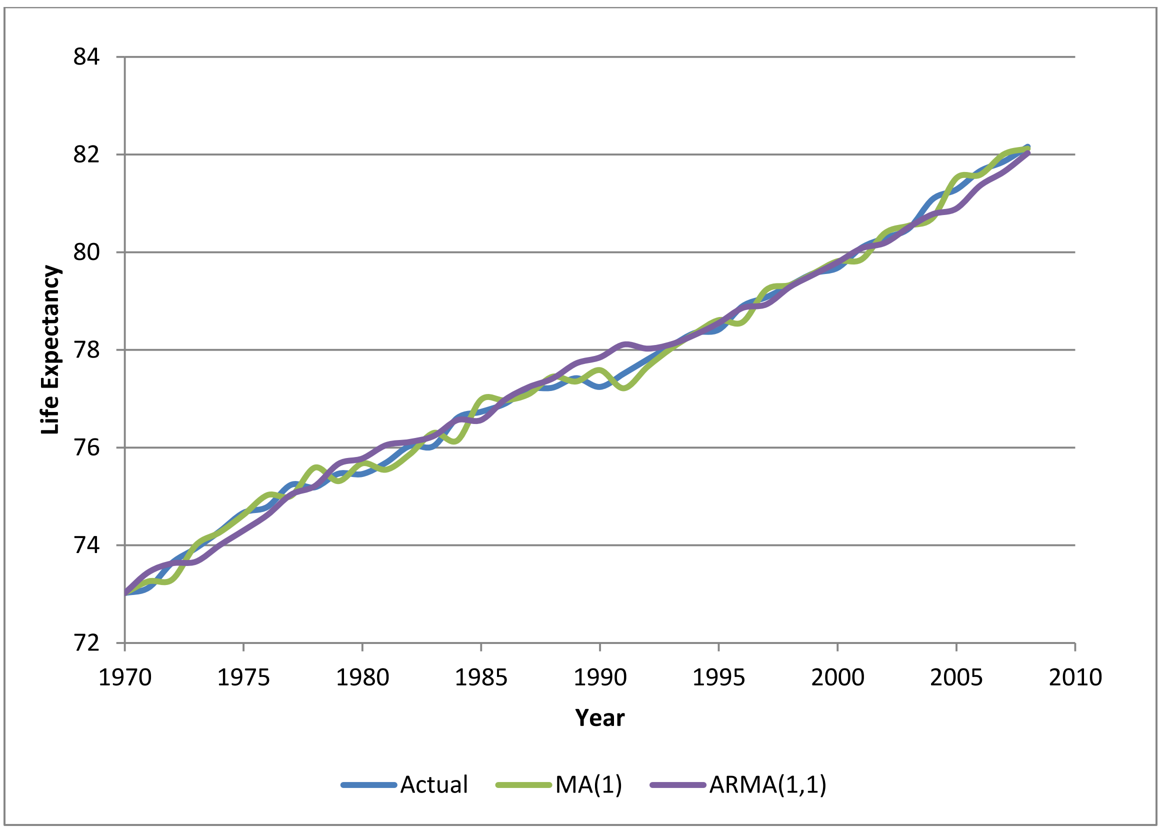
Modeled vs Actual

Since AR(1) failed the Box-Pierce test, I removed it from the analysis. Using the MA(1) and ARMA(1,1) models for forecasting, I calculated the forecasted first difference values in the supporting Excel spreadsheet, and the forecasted first difference values were plotted against the actual first differences.





For the most part, both models show fairly similar degrees of variability, and with a few exceptions, the residuals were contained within the -0.4 to 0.4 range.



Fitting both models to the actual life expectancy, it does appear that the ARMA(1,1) model traces the original series a little better, and the MA(1) line oscillates more about the actual life expectancy line.

Conclusion

With expansion of healthcare access and the improved quality of living, the life expectancy at birth has been steady rising over the years.

In analyzing the data, I note that while the actual series is not stationary, the first differences of the series is relatively stationary, and that is best modeled by an ARMA(1,1) process. Therefore, the original series is best described by an ARIMA(1,1,1) process, which makes intuitive sense, as we would expect last year's life expectancy to influence the predicted life expectancy.