September 19, 2000
date last saved: 09/19/00 2:46 PM
date last printed: 09/19/00 2:46 PM

Assessing the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality

Ronald Lee
Demography and Economics
University of California
2232 Piedmont Ave
Berkeley, CA 94720
rlee@demog.berkeley.edu

Timothy Miller
Demography
University of California
2232 Piedmont Ave
Berkeley, CA 94720
tmiller@demog.berkeley.edu

Research for this paper was funded by a grant from NIA, AG11761.We thank John Wilmoth for making available mortality data for the US, France, Sweden, and Japan, through the Berkeley Mortality Data Base. We thank Statistics Canada and Francois Nault for making Canadian data available. John Wilmoth and Ken Wachter provided very useful suggestions for the analysis. This paper is available in electronic form at www.demog.berkeley.edu.

Abstract

The Lee-Carter method for forecasting mortality was published eight years ago, with an application to US mortality data, 1900-1989. The method has been quite well received, but there have also been criticisms. Some have thought that the probability bands are implausibly narrow. Others have argued that many age specific rates are so low that they can't realistically be projected to decline much further. Some argue that it must be sub-optimal to ignore biomedical information that might inform the forecasts, and that forecasts based on expert opinion should be preferred. Some have called for more within sample testing of the methods, and others have questioned whether the ax and bx should be treated as invariant. Bell (1997) noted that the model did not fit the jump off data very well. In this paper we will examine many of these issues.

This paper will assess the performance of the 1992 forecast over the years since 1989. It will also conduct some more demanding tests of its performance within sample for the US as well as for some other countries. It will compare within sample performance to the performance of the projections of the Social Security Administration (SSA) and some other US forecasts. It will consider some extensions and modifications of the original procedure.

Results include:

- The original forecast started with an initial level of e0 that was .6 years higher than the actual for 1989. This error was carried over to all subsequent years of the forecast. Adjusting for this error in data for initial level, the forecast was within 0.2 years of e0 in 1998 and similarly close to the rates of decline of the individual age groups from 1989 to 1997.

- Applying the method retroactively to project to e0 in 1998, using only data available up to each historic start point, the hypothetical forecasts are quite accurate, with forecasts starting in 1946 having errors of two years or less. The 95% probability bounds contained the true value for 1998 85% of the time.

- We analyze 78 hypothetical forecasts with jump-off years from 1920 to 1997 and forecast horizons from 78 years to 1 year. The method tended to under-predict gains in life expectancy in the US, particularly when launched from earlier dates. 91% of errors at 31-40 year horizons were negative (predicted e0 less than actual) and 100% of errors beyond a 50 year horizon were negative. The true e(0) fell within the 95% probability interval for 2,984 out of 3,081 forecasted e(0) values or 97% of the time. The probability bounds appear to be too broad for horizons up to 40 years and too narrow for horizons beyond 50 years.

- The average error and mean squared error for LC forecasts since 1950 are substantially lower than those of SSA since 1950.

- If the method had been used to forecast 1995 e0 for Sweden, starting in 1950, it would have been right on target until 1980, and two years too low in 1995. Results

for France and Canada are very similar. For Japan, the data only start in 1950; forecasts from 1975 to 1996 are below the actual value, and one year too low by 1996. Looking at all the forecasts combined, the 95% probability bounds contain the actual e(0) values for 152 out of 162 forecasted values or 94% of the time.

- There have been very significant changes in the relative rates of decline of mortality by age, in the US, Sweden, France, Canada, and Japan, contrary to an assumption of the original method. This requires that the ax and bx coefficients be estimated on data since 1950 or so, not over the whole century.

- Forecasts should use actual last observed death rates as the base for forecasts, as described in the paper. Second stage fitting can be done more easily using actual e0 as the fit criterion in place of matching the total number of deaths.

I. Introduction

Important policy decisions are made today based on forecasts of the elderly population 75 years in the future. Pension policies are the prime example. Fundamental changes in the US Social Security Administration are under consideration in part because of a financial crisis for the system which is based on long term population projections. Old age dependency ratios are the key variable in these forecasts, and they depend on the number of elderly in the numerator, and the number of working age people in the denominator. The denominator depends heavily on future trends in fertility and perhaps migration, and these are notoriously difficulty to forecast. The elderly in the numerator have already been born, at least for forecasts over a 65 year horizon, and so they are on firmer ground. Yet the record of demographers and official agencies in forecasting their numbers is flawed. A series of studies by Keilman (1999; 1997) has found systematic under prediction of the elderly population in industrial nations, by about .5% for each year of a forecast, so that after 75 years one might expect the actual number to exceed the forecast by as much as 60%! (=1/(1-75*.005)). For the "oldest old", those over 85, the under-prediction occurs at about 1% per year of the forecast, so after 75 years the actual number could exceed the forecast by as much as 300% (=1/(1-75*.01)). While immigration must have contributed to these errors, the main culprit is the systematic under-prediction of mortality decline and life expectancy gain. We will suggest that these problems continue in the recent and current forecasts of industrial nations.

In this paper, we will present an ex post assessment of the performance of the mortality forecasts of the Social Security Administration, and find some evidence of bias towards the prediction of smaller increases in life expectancy at birth than subsequently occurred. Our main purpose, however, is to make a careful and detailed assessment of the performance of the Lee-Carter method for forecasting mortality. We evaluate the performance both in terms of projected e(0) and projected age-specific mortality rates. We first examine the performance of the forecasts published in 1992 relative to subsequent mortality trends. We next construct forecasts with jump-off years earlier in the 20th century, pretending we had only the data available up to that point, and comparing the subsequent forecasts to the actual outcomes. We also conduct some similar, but less detailed, experiments using the method to produce forecasts for Japan, Canada, France and Sweden, with jump-off year in 1950. And last, we examine age patterns of decline during the 20th century and consider the possibility that the age pattern has changed over time contrary to the assumptions of the method.

II. Original article (1992)

A. Overview of the LC approach

Lee and Carter (1992, henceforth LC) developed a new method for modeling and forecasting mortality, and used it to forecast US mortality to 2065. Since that time, the method has attracted a certain amount of attention. The most recent Census Bureau population forecasts (Hollmann et al., 2000) use the Lee-Carter forecast as a benchmark for their long-run forecast of US life expectancy. The two most recent Social Security Advisory Panels have recommended the adoption of the method, or forecasts consistent with it, by the Trustees. The method has also been applied in a number of other countries (most recently for the G7 nations, see Tuljapurkar et al., 2000). We begin this section with a brief overview of the approach followed by an assessment of the performance of the 1992 forecast over the years since 1989, the jump-off year of the forecast.

The basic LC model of age specific death rates (ASDRs, and denoted $m_{x,t}$) is:

$$\ln\left(m_{x,t}\right)=a_x+b_xk_t+\varepsilon_{x,t} \qquad \text{(Equation 1)}$$

Here a_x describes the general age shape of the ASDRs, while k_t is an index of the general level of mortality. The b_x coefficients describe the tendency of mortality at age x to change when the general level of mortality (k_t) changes. When b_x is large for some x, then the death rate at age x varies a lot when the general level of mortality changes (as with x=0 for infant mortality, for example) and when b_x is small, then that death rates at that age vary little when the general level of mortality changes (as is often the case with mortality at older ages). Note that the model assumes that all the ASDR move up or down together, although not necessarily by the same amounts, since all are driven by the same period index, k_t. In principle, not all the b_x need have the same sign, in which case movement in opposite directions could occur, but in practice, all the b_x do have the same sign, at least when the model is fit over fairly long periods. Note that the proportional rate of decline of any death rate is give by $b_x(dk/dt)$. If dk/dt is constant, that is if k_t is declining linearly, then each ASDR will decline at its own age specific exponential rate, proportional to b_x, and depending on the rapidity of the decline in k_t. The same model was selected by Gomez de Leon (1990) using exploratory data analysis on the historical data for Norway, out of a larger set of possibilities.

The strategy is to estimate this model on the historical data for the population in question, obtaining values for a_x, b_x and k_t. The values of k_t form a time series, with one value for each year of data. Standard statistical methods can then be used to model and forecast this time series. LC selected a random walk with drift as the appropriate model, which has the form:

$$k_t=k_{t-1}+c+e_t \qquad \text{(Equation 2)}$$

In this specification, c is the drift term, and k is forecast to decline linearly with increments of c, while deviations from this path, e_t, are permanently incorporated in the trajectory. The variance of e_t is used to calculate the uncertainty in forecasting k over any given horizon. The drift term, c, is also estimated with uncertainty, and the standard error of its estimate can be used to form a more complete measure of the uncertainty in forecasting k.

The projected k can then be used in Equation 1, together with the estimated a_x and b_x, to calculate forecasts of the ASDRs, and from these any desired life table functions can be derived. The probability intervals on the forecasts of k can then be used in the same way to calculate intervals for the forecasts of the ASDRs, and (because these are all linear functions of the same k) the forecast of e0. However, forecast errors in the ASDRs and e0 derive additionally from the $\varepsilon_{x,t}$ and from uncertainty about the true values of a_x and b_x. LC show that these latter sources of error matter less and less as the forecast horizon lengthens, and they are dominated by uncertainty about k in the long run. For a forecast horizon of 10 years, 98% of the standard error of the forecast of e_0 is accounted for by uncertainty in k; for the individual age specific rates, the other sources of uncertainty are more important initially and remain important longer, but after 25 years most account for less than 10% of the standard error of the forecasts (see LC table B2).

From inspection of Equation 1 it is apparent that there is no observed variable on the right hand side of the equation, so ordinary regression methods cannot be used to estimate the model. LC describes a simple approximate method using regression methods, but the Singular Value Decomposition (SVD) gives an exact least squares fit. Also note that if a_x, b_x and k_t is one set of coefficients for the model, then a_x, b_x/A and A*k_t will be an exactly equivalent set, for any constant A. Similarly, $a_x - b_x$*A, b_x, k_t(1+A) will also be an equivalent formulation for arbitrary constant A. LC stipulated a unique representation by setting a_x equal to the average of the logarithms of mx,t over the data period, and setting the average value of k_t equal to zero. In this case the sum of the b_x values is unity.

The method has a number of appealing features. The basic model is very simple, and although its use for forecasting involves a number of steps, each is simple in itself. The method is “relational” in demographers’ terminology. That is, it involves the transformation of actual existing mortality schedules for each study population, and therefore on the one hand is largely non-parametric, and on the other hand incorporates particular features of the mortality pattern of a given population. The method is also probabilistic, in the sense that it involves statistical fitting of models, and the quality of the fit of the historical data can be used to provide probability intervals for the forecasts. As a matter of empirical fact, in the applications of the method to date, involving at least ten national data sets, the historical trend in k has always been found to be highly linear with time, and the random walk with drift has been found to give a good fit. This approximate linearity is useful for forecasting. It contrasts with the typically nonlinear trajectories of life expectancy, which rises at a decelerating rate when age specific mortality rates decline at constant exponential rates. Finally, the method can also be used as the basis of a simple model life table system, and indirect estimation methods can be developed to expand the mortality data available as the basis for forecasting.

B. Assessing the original forecast

In their original article, LC noted that the model would not fit the age specific mortality data exactly in the jump off year, which would mean that the initial conditions for the forecast would not be quite right. This would inevitably lead to error which would be particularly important in the early years of the forecast. They noted that it would be possible to set a_x equal to the most recently observed log age specific rates, and thereby fit the initial conditions exactly (with k_t = 0). However, they argued that this practice might extrapolate idiosyncratic features of mortality in the jump off year, and it was therefore preferable to estimate a_x as the average values of the log death rates (LC:665-666). In retrospect, this appears to have been a mistake, since the error in e_0 of .6 years at the jump off year caused significant bias in the forecasts for the first decade, as we shall see below, and as Bell (1997) has pointed out (LC estimated e_0 for 1989 at 75.66 years, whereas official data puts it at 75.08). Bell (1997) assessed the performance of four mortality forecasts: LC (as published); LC (with the jump off year corrected); McNown-Rogers; and the SSA actuaries. He concluded that the LC forecasts did better than the SSA or McNown-Rogers, but that a corrected LC forecast did better still.

Figure 1 displays the original LC mean forecast of e0, a similar forecast but with the correct jump-off level, and the SSA projections done at the same time. The bias in the original LC projections is clearly apparent, but it is also apparent that those projections correctly identified the trend in e_0. SSA appears to be somewhat low, ending up about 0.8 years below the actual e0. The adjusted LC is about 0.2 years too low in 1998 (the latest data available to us). Over this period, the actual e0 always remains well within the 95% prediction interval for both the original LC and the adjusted LC.

If the forecasts of e_0 performed well from 1989 to 1998, how about the forecasts of the individual age specific rates? Once again, there are certainly errors due to the errors in initial conditions. Figure 2 instead focuses on the LC projected age specific rate of decline of death rates from 1989 to 1997 for sexes combined, since this will not be affected by the errors in initial rates. It also plots the actual rates of decline, and those projected by SSA. The agreement between the LC forecast and the actual rates of decline is striking, particularly at the older ages. The SSA projections, however, incorrectly forecast slower mortality decline in the young adult years. We will return to this topic later, for a different perspective on the age pattern of decline.

C. Criticisms and advances since publication

The method has been quite well received, but there have also been criticisms. Some have thought that the probability bands are implausibly narrow (e.g. Alho, 1992:673). Others have argued that many age specific rates are so low that they can’t realistically be projected to decline much further. Some argue that biomedical information should inform the forecasts, perhaps through incorporating expert opinion as is done by the Social Security Actuaries. Some have called for more within-sample testing of the methods, and others have questioned whether the a_x and b_x should be treated as invariant. Bell (1997) noted that the model did not fit the jump off data very well. In this paper we will examine many of these issues.

Considerable work has been done to refine and extend the method since the original LC article. Wilmoth (1993) has developed improved fitting methods based on weighted least squares. Methods for modeling and forecasting regional systems of mortality have been developed (Lee and Nault, 1993). Better procedures for dealing with the jump-off year have been developed (Bell, 1997). Alternatives for modeling mortality for the oldest old have been explored. Consideration has been given to the special role of leader and follower countries (Wilmoth, 1998). The method has been applied to cause of death data (Wilmoth, 1998) to sexes separately, and by race. (Carter and Lee, 1992; Carter 1996). There have been many applications to countries other than the US (e.g., Lee and Rofman, 1992; Tuljapurkar et al., 2000). Lee (2000) provides a summary of the model's development, extensions, and applications such as stochastic forecasts of social security system finances.

III. Assessing LC on US time series, within sample

A. The nature of the tests

In the original LC article, there were some tests of forecast performance within the historical data period, but none of these involved re-estimating a_x, b_x and k_t. Instead, time series models were fit to different portions of the time series of estimated k_t. Here we will make a more rigorous test, in which we refit the model from scratch on each chosen sub-sample of data. Our earliest experimental forecast is based on data from 1900 through 1920. Our next uses data 1900 through 1921; our next through 1922; and so on until our last forecast uses data from 1900 through 1997 to make a forecast for 1998. In this way, we have 78 different forecasts for mortality one year ahead; 77 for a two year horizon; and finally one with a 78 year horizon. We re-estimated the a_x and b_x for each set of data, and then re-estimated k_t for these years conditional on these a_x and b_x estimates, by choosing k_t (in the second stage) so as to match exactly the given value of e_0 in the data for that year.[i] This departs slightly from the procedure in the original LC, where k_t was chosen to match total deaths, which requires annual age-distributed population data as well.

Once k_t was estimated for each year of the sample, we did not carry out standard diagnostic methods to choose an optimal ARIMA model for each data sub-sample, but rather assumed that the random walk with drift model held. It was fitted and used to forecast k_t over the desired time range.

LC introduced a dummy variable for the influenza epidemic of 1918. Our preference today is to include the dummy (permitting a one time positive change in k in 1918, followed by a one time equal negative change in k in 1919), and in the forecast to incorporate a 1/T chance of an identical positive and negative change in k occurring, where T is the length of the base period over which the model was fit. This has a small effect on both the mean and the variance of the forecast. We did not do this for these experimental forecasts, here described.

B. Forecasting to 1998 (e0)

Figure 3 plots all 78 forecasts for life expectancy in the year 1998, each from a different jump-off year, and each over a different forecast horizon. Each forecast for 1998 is plotted above its jump-off date. The 95% probability intervals are also plotted. The horizontal line indicates the observed value of life expectancy for 1998, so it is the true value relative to which the forecasts can be assessed. There are several points to note. *First*, although the experimental forecasts tend to be too low, they are generally fairly close to the actual value for 1998. The earlier forecasts, using data up through the 1920s and 1930s are on average five years below the true value; beginning in 1946 forecasts are within two years of the correct value. Over all, the mean forecasts look quite good. *Second*, the 95% probability intervals failed to contain the true value for 1998 in 12 out of the 78 forecasts, or 15% of the time, compared to the 5% which was intended. *Third*, the median forecast for 1998 fell below the actual value for 1998 in 74 of the 78 forecasts, or 95% of the time.

C. Errors by forecast horizon (e0)

It is also useful to assess forecast errors (forecast-actual) by horizon. We have done this for horizons of 1, 5, 10, 20, 40 and 60 years. For a 1 year horizon, we have 78 different jump-off dates, while for the 60 year horizon, we have only 19. For each forecast, we find the percentile in its probability distribution where the observed value falls. For example, if the actual corresponds to the median of the forecast distribution, we assign it 50. If it corresponds to the lower 7% of the distribution, we assign it 7; and so on. We then plot the frequency distribution of these percentile scores. If the probability distribution associated with each forecast does in fact describe the probability distribution of errors, then this frequency distribution should be uniform between 0 and 100. If the actual distribution of percentiles is more concentrated in the middle, around 50, that indicates that the distribution of the errors is more tightly clustered then our forecast leads us to expect, and if there are less in the middle of the distribution and more towards the 0 and 100 end, then our forecast understates the width of the error distribution. If most of the true values fall below the 50^{th} percentile, then most of the time we have over-estimated. While if they fall above the 50^{th} percentile, then we tend to systematically underestimate the true value.

Figure 4 plots the histogram of the percentiles for each horizon.

Table 1 presents various measures of forecast performance, including the Mean Squared Error (MSE), the Mean Absolute Percent Error (MAPE), the average error (Bias), the percent of positive errors, and the proportion of actual values that fall within the 95% probability interval of the forecast. The table reports performance by forecast horizons as well as a summary over all forecast horizons.

Table 1

Forecast Horizon	Average error	MAD	RMSE	MAPE	Number of estimates	% under-projected	% within 95% probability interval
1-5	-0.11	0.45	0.60	0.16	380	54	99
6-10	-0.32	0.82	1.03	0.47	355	56	100
11-20	-0.73	1.23	1.60	1.15	635	67	97
21-30	-1.37	1.47	1.99	2.03	535	84	100
31-40	-1.68	1.73	2.14	2.45	435	91	100
41-50	-2.23	2.25	2.75	3.41	335	96	95
51-60	-3.54	3.54	3.75	5.07	235	100	89
61-78	-4.38	4.38	4.53	5.39	171	100	80
ALL	-1.49	1.76	2.34	2.45	3,081	78%	97%

The method tended to under-predict gains in life expectancy in the US, particularly when launched from earlier dates. 91% of errors over 31-40 year horizons were negative (predicted e0 less than actual) and 100% of errors beyond a 50 year horizon were negative. The 95% confidence bounds contain the actual e(0) value 97% of the time. But, they appear to be too broad for intervals up to a 40 year horizon and too narrow for those beyond a 50 year horizon.

D. Error correlations by age, horizon

As noted briefly above, Equation 1 has an error term, $\varepsilon_{x,t}$, since the expression does not provide a perfect representation of variation in age specific rates over time. In formulating the probability intervals for the life expectancy forecasts, this error term was ignored, and only errors arising from the innovation in k_t and from errors in estimating the drift term, were incorporated. If we were interested only in e(0) and if the $\varepsilon_{x,t}$ were uncorrelated across age, this assumption might be relatively harmless, because some twenty different values of $\varepsilon_{x,t}$ enter into the calculation of any life expectancy, and the average effect should be very small. However, if the errors are correlated, such that those for older ages tend to move together and those for younger ages tend to move together, then they might have an important influence even on life expectancy. There are also errors in the estimation of the a_x and b_x coefficients, which are not taken into account in our probability intervals for the e_0 forecasts.

In general, we find that forecast errors tend to be strongly correlated at younger ages, less so at older ages, and young errors are only weakly correlated with errors at older ages. At longer horizons, correlations become more positive due to dominance of errors in k. Figure 5 provides some examples for select age groups and forecast horizons. Further work on the analysis of age-specific errors is underway.

IV. Assessing LC on historical time series from other countries

We also carried out within sample tests for Sweden, Japan, France and Canada. The results are shown in the panels of Figure 6. For France, where both WWI and WWII had profound effects on mortality, we have dummied the effects in a similar way, but not allowed for a possible recurrence in the future. Allowing for a recurrence would greatly increase the variance of the forecast. Such decisions reflect the judgment of the analyst.

If the method had been used to forecast 1995 e0 for Sweden, starting in 1950, it would have been right on target until 1980, and two years too low in 1995. Results for France and Canada are very similar. For Japan, the data only start in 1950; forecasts from 1975 to 1996 are below the actual value, and one year too low by 1996. Looking at all the forecasts combined, the 95% probability bounds contain the actual e(0) values for 152 out of 162 forecasted values or 94% of the time.

V. Changing age-shape of mortality

A number of people have suggested that the b_x coefficients might vary over time; this possibility was not explored by LC. Kannisto et al. (1994) found that the rate of mortality decline had been accelerating over recent decades for ages 80 to 100. Horiuchi and Wilmoth (1995) show that in a number of countries, mortality declines at older ages now take place more rapidly then at lower ages, reversing the historical pattern. This research suggests that it is important to take very seriously the possibility that the age pattern of mortality decline may alter over time, and may not be well described by a fixed set of b_x coefficients. Note that the a_x coefficients will always be changing over different historical periods, because they are the average log death rates, and these averages will change in level as mortality falls, and change in shape because the b_x coefficients tell us that at different ages, mortality declines at different rates. This poses no problem, because the changing shape and level of the a_x are implicit in the b_x, and no additional treatment is necessary.

Recall that our earlier examination of the post-publication performance of LC showed that it correctly forecast the age pattern of mortality decline as well as the increase in e_0 over the past 9 years. This suggests that the fixed b_x assumption has worked well. However, a closer examination of the age pattern of decline in the US shows otherwise. Figure 7 plots the average rate of decline for sexes combined mortality by age for 1900 to 1949 and for 1950 to 1995. It is clear that there has been an important change, with mortality now declining at roughly the same rate across all ages above 15, whereas for the first half of the century it declined far more rapidly at the younger ages.

Examination of the historical pattern of decline in Japan, Sweden, Canada, and France shows similarly striking changes, with a flattening of the age profile of decline. (See Figure 8).

Is this a long term change, routed in the changing cause structure of mortality, or in the resistance of mortality at different ages to biomedical progress? Or is it due to what we might hope will be more transitory influences on young adult mortality in industrial nations, such as AIDS and accidents? We are not sure. But the more prudent course is to assume that these changes are long term, and to incorporate them into our forecasts in one way or another. A simple and satisfactory solution, adopted by Tuljapurkar et al. (2000), is to base the forecast on data since 1950, and assume fixed b_x over that range but not over the whole century.

VI. Comparison of official forecasts from SSA and others to LC forecasts

A. Forecasting to 1998

We have examined the historical record of SSA projections, including two earlier ones that were used by SSA but prepared by other agencies. Figures 9 and 10 examine forecasts of e(0) for the year 1998. Figure 9 compares the middle series forecast from SSA with the median LC forecast. The figure shows that the official projections have been systematically too low – by 12 years in 1930, about 7 years in the 1940s, then by 2 to 4 years until those done in 1980, which then jumped to being too high. It can be seen that the SSA estimates reacted strongly to the slow mortality gains of the 1960s, and then to the rapid gains of the 1980s. By contrast, the LC method responds only modestly to these fluctuations, since they only modestly affect the average trend over the century. The LC method also tends to be somewhat low in early years, but performs substantially better than SSA. It would have been closer to the true value in 1998 for most forecasts. It picks up the correct track for 1998 considerably earlier.

Figure 10 shows the high-low range of SSA projections along with the 95% probability interval of LC. The true value of e(0) for 1998 lies beyond the high bound for most of the SSA forecasts up until 1970.

B. Errors by horizon, comparison to LC

In assessing errors by forecast horizon, we have restricted our sample to post-1950 government forecasts. We have only 3 early government forecasts (pre-1950) – which provided e(0) forecasts for only a few select years in the future. This makes the analysis of errors by length of horizon complicated for these groups. We are working on obtaining more of the data for these early forecasts. For comparison to LC, we use both the full sample (1920-1997) and a restricted sample which matches the time period of the SSA forecasts (1950-1997). For LC, we have forecasts for every year. For SSA, the forecasts are issued irregularly. In our calculations we have weighted each SSA forecasts by the reciprocal of the number of forecasts issued within the decade. In this way, each decade contributes equally to the error estimates.

Figure 11 compares the average bias in the SSA and LC forecasts by length of forecast horizon. Horizons are by single year from 1 to 7 and then grouped (8-12, 13-17, 18-22, 23-27, 28-38, 39-46, and 39-60 years). SSA forecasts issued since 1950 compare favorably with LC forecasts issued since 1920. However, when we examine those LC forecasts issued during the same time period (since 1950), we find that LC performs substantially better.

Figure 12 compares the root mean square error (RMSE) for SSA and LC forecasts. SSA forecasts perform slightly better than those of LC for the first and second years. At all horizons beyond 2 years, LC performs better than SSA and substantially better as the forecast horizon increases.

C. General problem of official forecasts

Government forecasts generally rely on expert opinion for their long-run forecast. The evidence suggests that this has resulted in forecasts which are too pessimistic. The early reports were issued during the Great Depression and the Second World War. Perhaps these events influenced expert opinion about future progress. And yet, at that time, the data were telling a different story, since mortality had been declining quite rapidly over the previous decades. A quote from the 1943 report is interesting in this regard. Thompson and Whelpton state their objection to statistical forecasting methods such as extrapolation: "More important, the extrapolation of past trends according to such formulas might show future trends which seemed incompatible with present knowledge regarding the causes of death and the means of controlling them." (National Resources Planning Board, 1943, p. 10). This suggests an alternative explanation for the pessimism of experts: present knowledge informs us about current limits, but not the future means of overcoming them. The Lee-Carter approach bases its long-run forecasts on the century-long decline in mortality in which limits have been continuously confronted and overcome.

VII. Conclusions

1) Lee-Carter (LC) forecasts of life expectancy and the age pattern of mortality performed quite well for the period since publication, at least after adjusting for an error in jump-off level.

2) Historical LC projections from various jump-off dates in the 20th century would have preformed well. For forecasts with jump-off after 1945, we are always within 2 years of the actual e(0) in 1998. The forecasts tend to under-predict future gains, especially those in the distant future. The 95% probability bounds contain the true value of e(0) 97% of the time. But, the bounds appear to be too broad for horizons up to 40 years and too narrow for those beyond 50 years.

3) Social security projections also have systematically under-predicted gains in e(0) since 1950. The average error and mean squared error for LC forecasts since 1950 are substantially lower than those of SSA since 1950.

4) LC life expectancy forecasts for Canada, Sweden and France with jump off year 1950 and for Japan with jump off year 1973 would have performed very well. But, like the US, would have systematically under-predicted actual gains.

5) Contrary to a basic assumption in the Lee-Carter model, the age pattern of mortality decline has shifted systematically in the US, Sweden, France, Canada, and Japan in the second half of the 20th century, with a flattening of the age specific rates of decline above age 15.

VIII. References

Alho, Juha M. (1992)

Bell, William R. (1997) “Comparing and Assessing Time Series Methods for Forecasting Age Specific Demographic Rates” *Journal of Official Statistics* 13:279-303.

Carter, Lawrence (1996) “Long-Run Relationships in Differential U.S. Mortality Forecasts by Race and Gender: Non-Cointegrated Time Series Comparisons”, 1996 Annual Meetings of the Population Association of America, New Orleans, May 9-11, 1996.

Carter, Lawrence and Ronald D. Lee (1992) "Modeling and Forecasting U.S. Mortality: Differentials in Life Expectancy by Sex," in Dennis Ahlburg and Kenneth Land, eds, *Population Forecasting*, a Special Issue of the *International Journal of Forecasting*, v.8, n.3 (November) pp.393-412.

Gomez de Leon, Jose (1990) Empirical DEA Models to Fit and Project Time Series of Age-Specific Mortality Rates," Unpublished manuscript of the Central Bureau of Statistics, Norway (July).

Hollmann, Frederick W.; Mulder, Tammany J.; and Kallan, Jeffrey E. (2000) “Methodology and assumptions for the population projections of the United States: 1999 to 2100,” Population Division Working Paper No. 38, U.S. Bureau of the Census.

Horiuchi, Shiro and John R. Wilmoth (1995) “The aging of mortality decline.” Presented at the Annual Meeting of the Population Association of America, San Francisco, April 6-8, 1995.

Kannisto, Vaino; Lauristsen, Jens; Thatcher, A. Rodger; and Vaupel, James W. (1994) “Reductions in mortality at advanced ages: Several decades of evidence from 27 countries.” Population and Development Review, Vol. 20, No 4, pp. 793-810.

Keilman, Nico (1997)

Keilman, Nico (1999)

Lee, Ronald D. (2000) “The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications,” *North American Actuarial Journal*, Vol. 4, No. 1, pp. 80-91.

Lee, Ronald D. and Lawrence Carter (1992) "Modeling and Forecasting the Time Series of U.S. Mortality," *Journal of the American Statistical Association* v.87 n.419 (September) pp.659-671.

Lee, Ronald D. and Francois Nault (1993) "Modeling and Forecasting Provincial Mortality in Canada," paper presented at the World Congress of the International Union for the Scientific Study of Population, Montreal, 1993.

Lee, Ronald D. and Rafael Rofman (1994) "Modelacion y Proyeccion de la Mortalidad en Chile," NOTAS 22, no 59, pp. 182-213. Also available in English from the authors, titled "Modeling and Forecasting Mortality in Chile."

Tuljapurkar, Shripad; Nan Li and Carl Boe (2000) "A universal pattern of mortality decline in the G-7 countries" *Nature*, in press.

Wilmoth, John R. (1993) "Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change," Technical Report, Department of Demography, University of California, Berkeley.

Wilmoth, John R. (1998) "Is the pace of Japanese mortality decline converging toward international trends?" *Population and Development Review* 24(3): 593-600.

---

[i] In all cases, the data we use are taken from the SSA data base, as maintained on the Berkeley Mortality Data Base web site, www.demog.berkeley.edu. The original LC article used NCHS data, and for the period before 1933 estimated age specific mortality and e0 indirectly using the age distribution of the total population, total deaths per year, and the ax and bx coefficients as estimated by SVD from the data 1933 to 1987, after the death registration area was complete.

Figure 1: Forecasts of life expectancy from 1989.

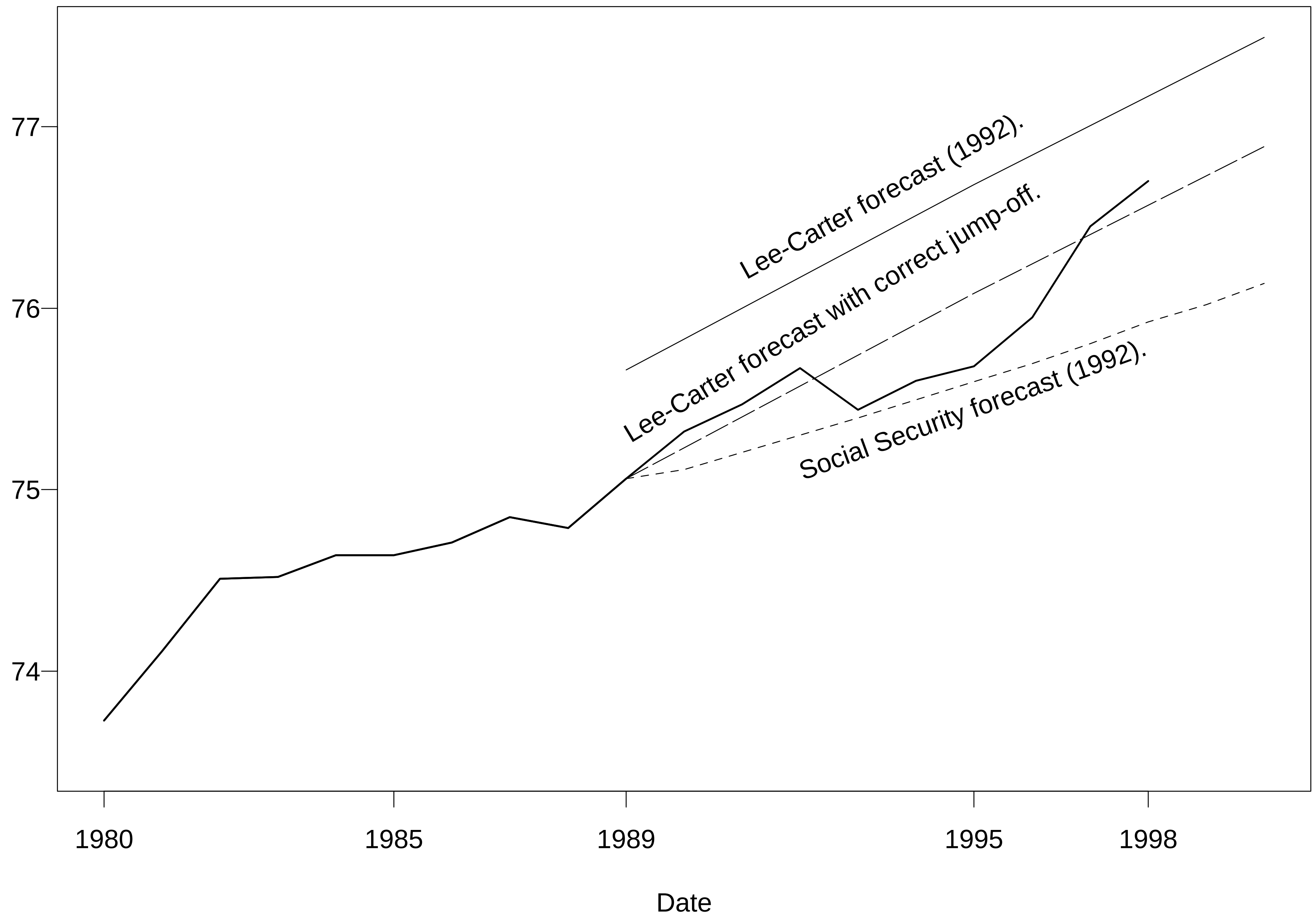

Figure 2: Average Annual Decline in Age-Specific Mortality, 1989-1997
Actual and Forecasts of Lee-Carter (1992) and SSA (1992)

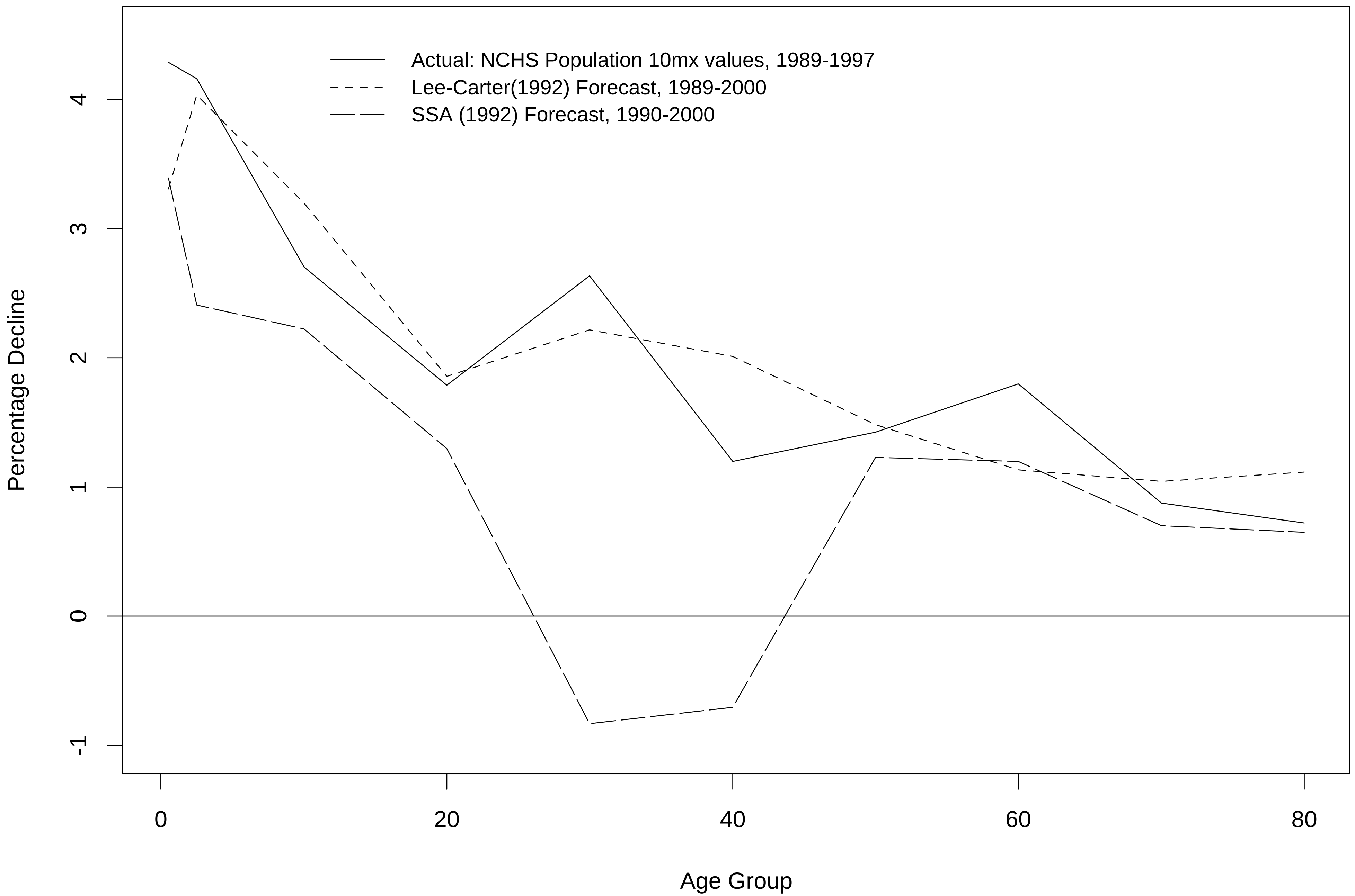

Figure 3: e(0) Forecasts for the Year 1998 by Forecast Date

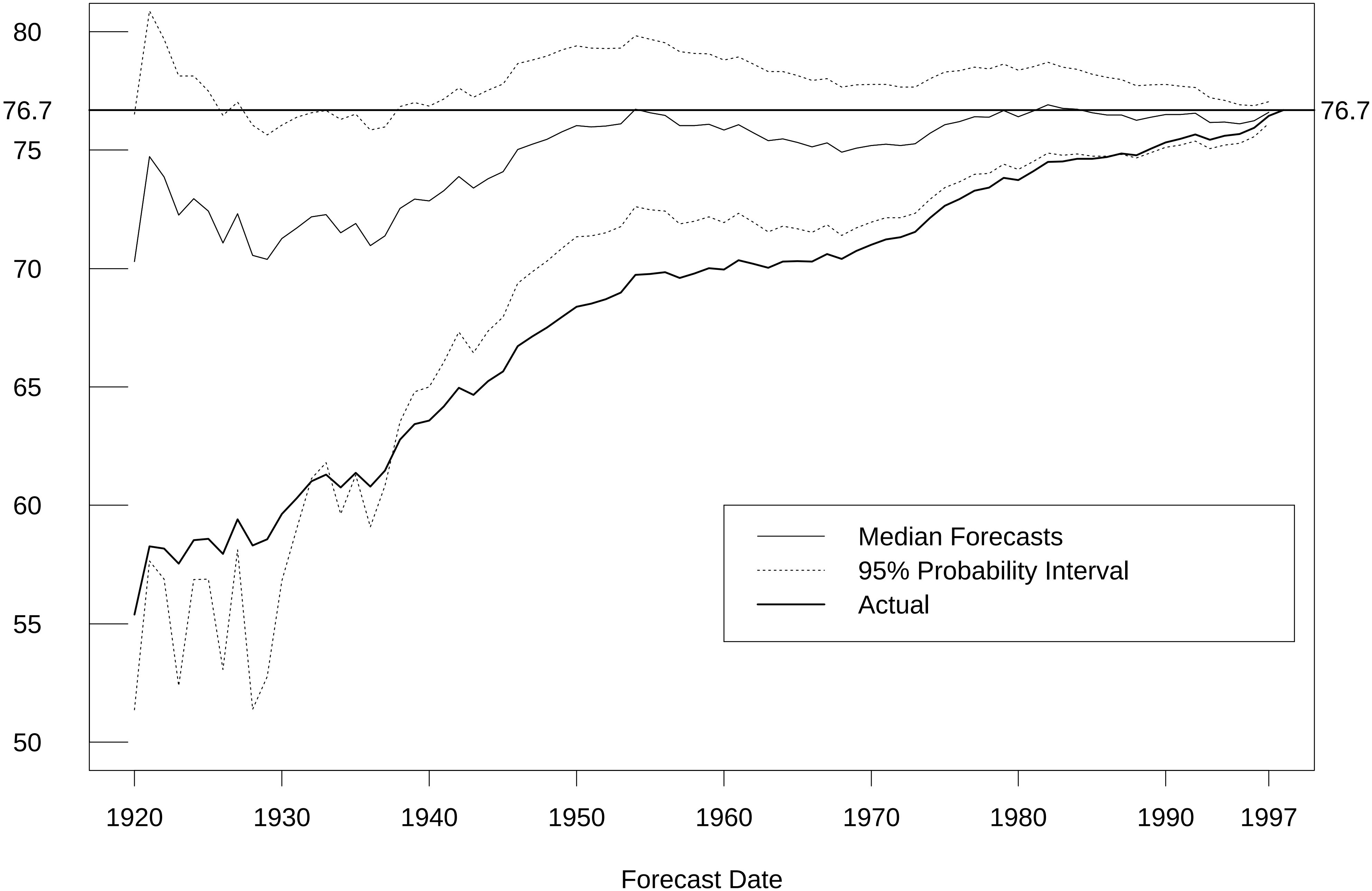

Figure 4: Percentile Error Distribution by Forecast Length

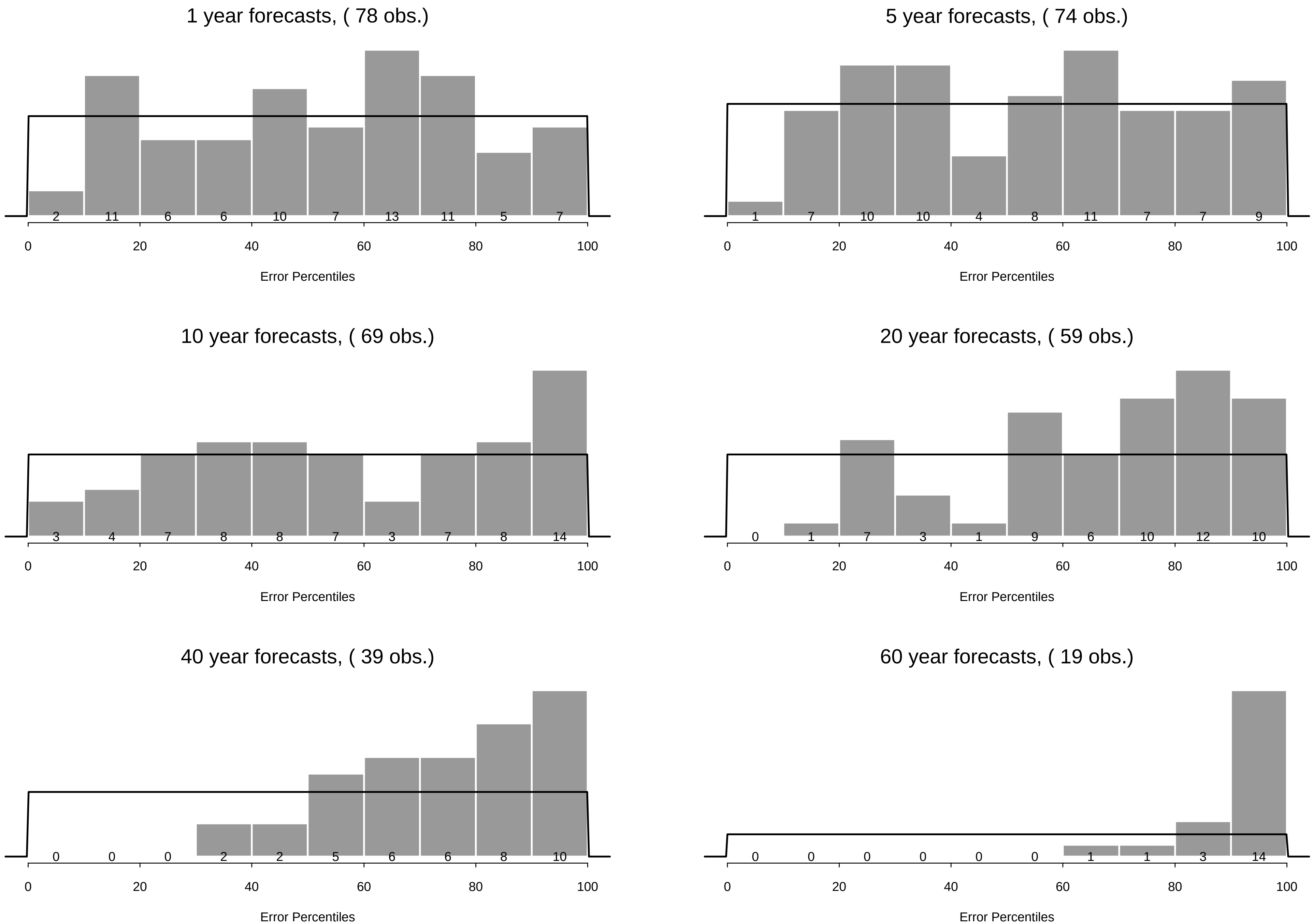

Figure 5: Error Correlations

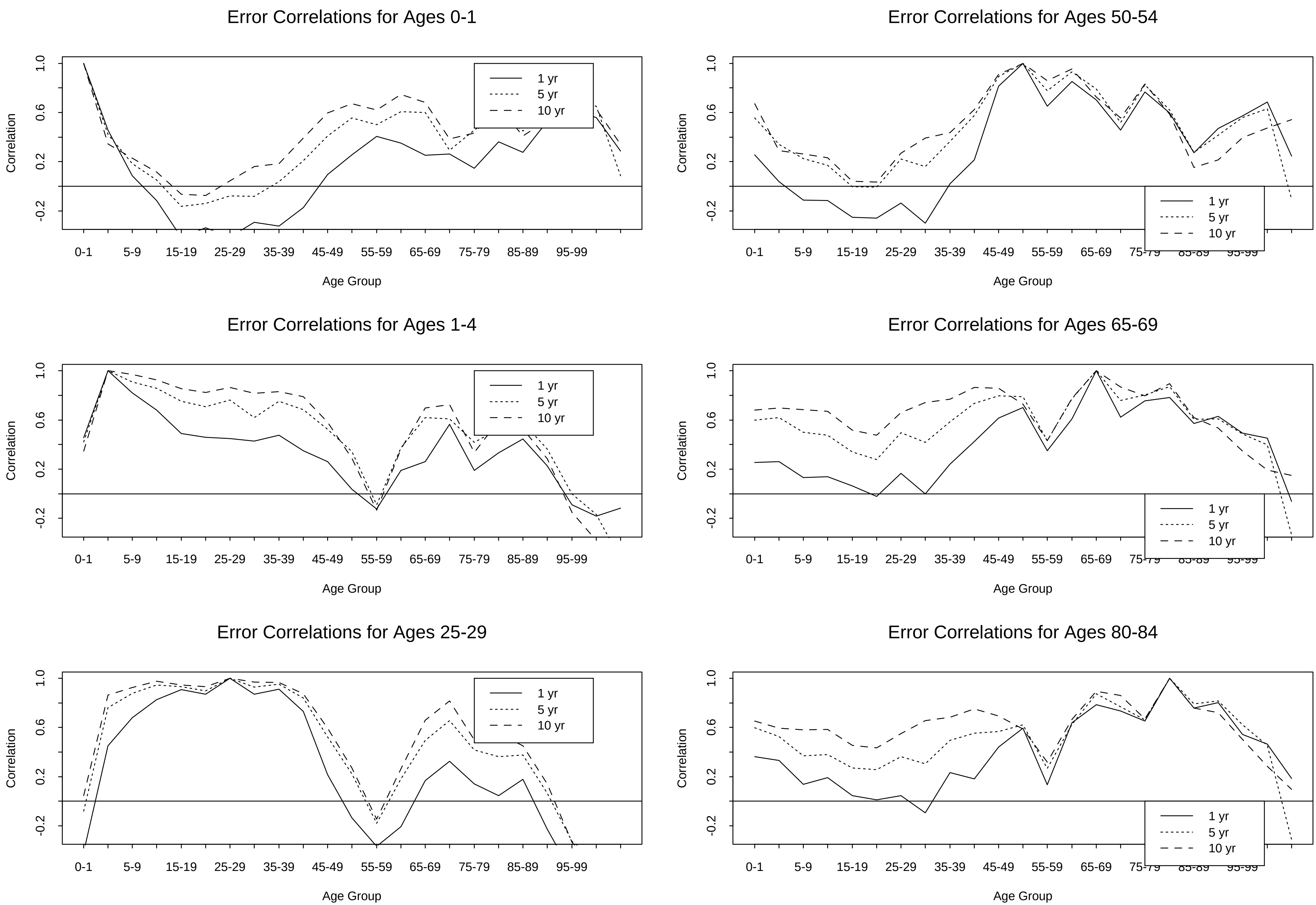

Figure 6: LC forecasts of life expectancy

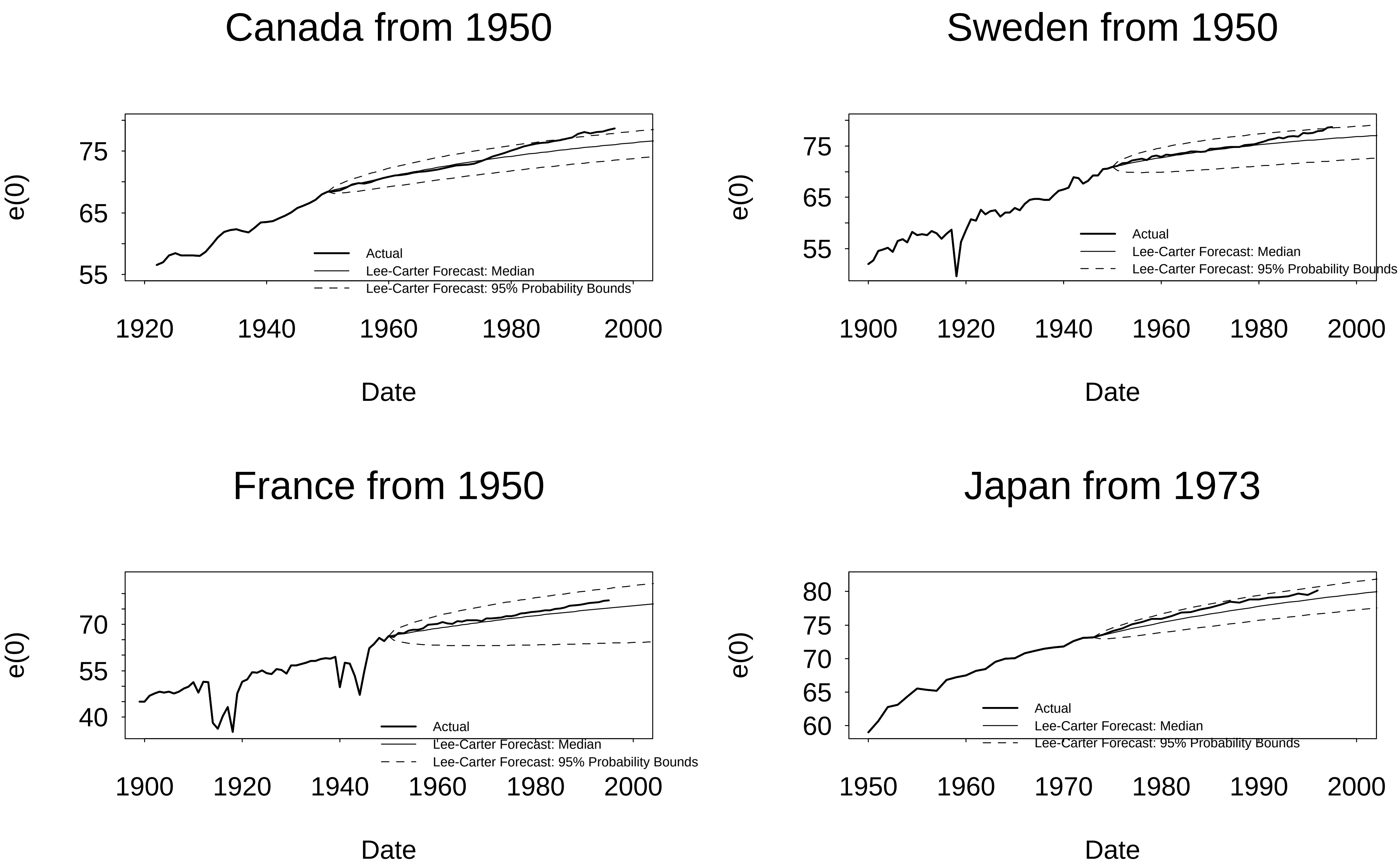

Figure 7: Average Annual Reduction in Age-Specific Death Rates, US

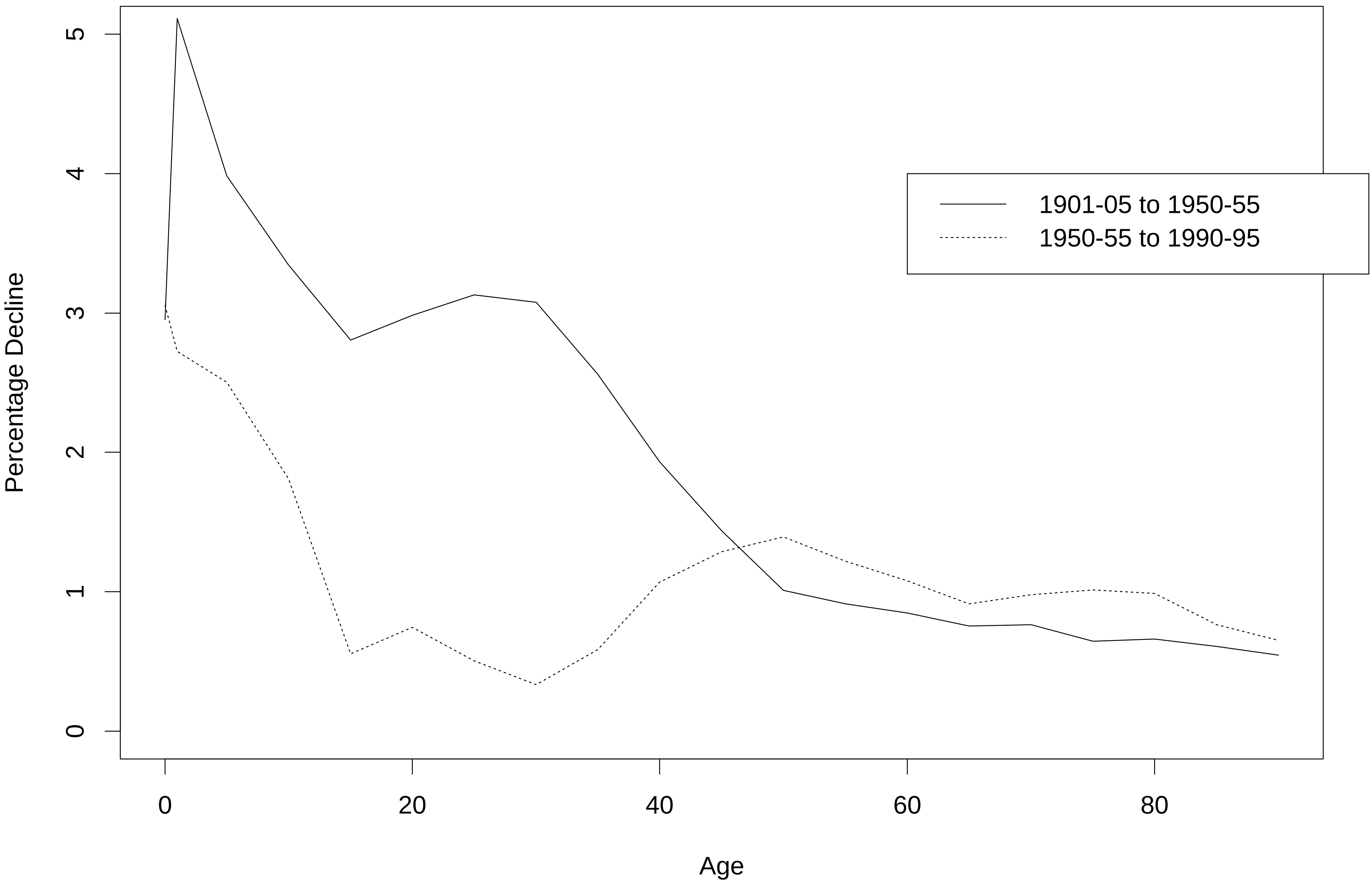

Figure 8: Average Annual Reduction in Age-Specific Death Rates

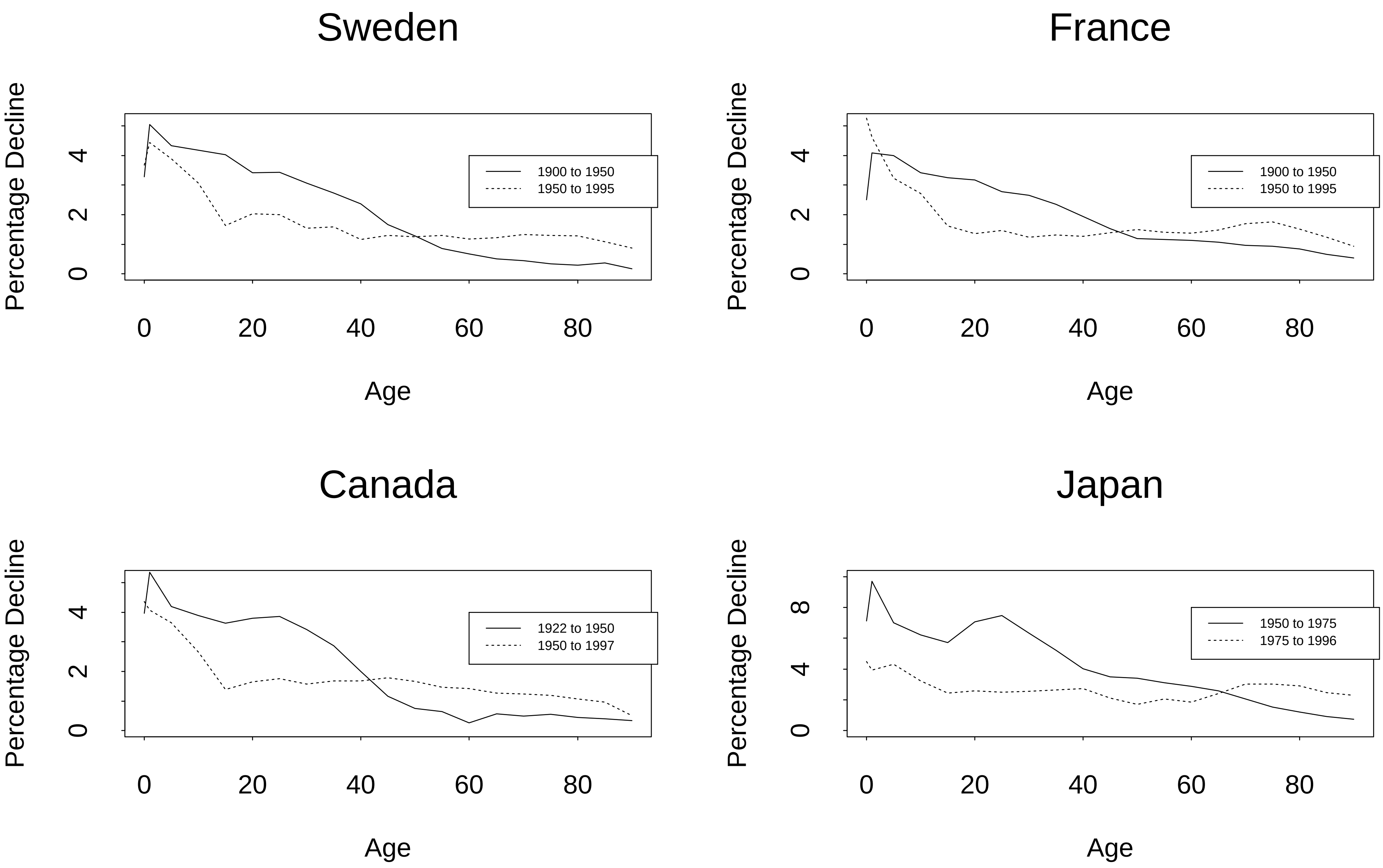

Figure 9: LC and SSA e(0) Forecast for 1998, by Forecast Date

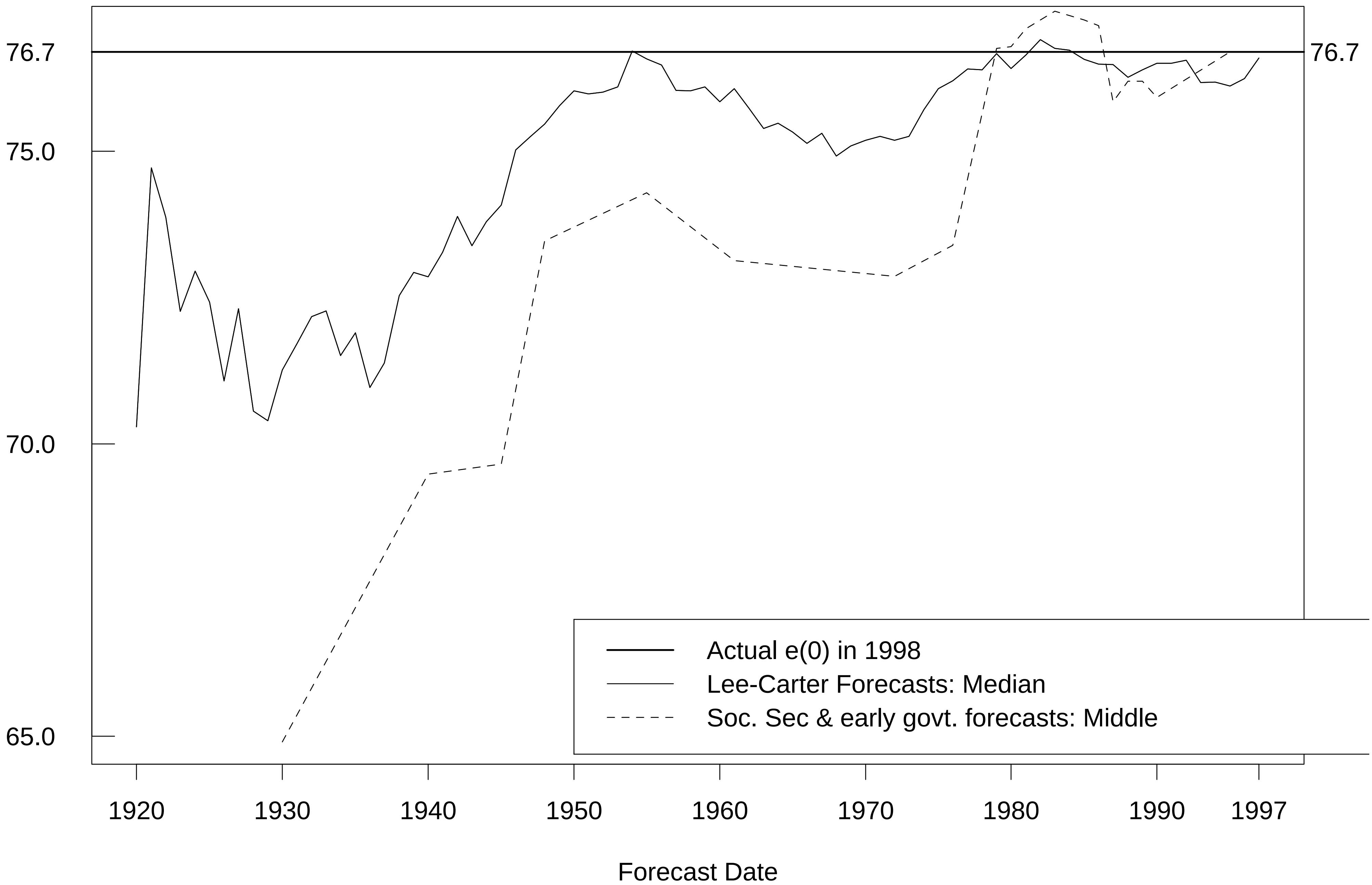

Figure 10: 95% Probability Interval and High-Low Range by Forecast Date

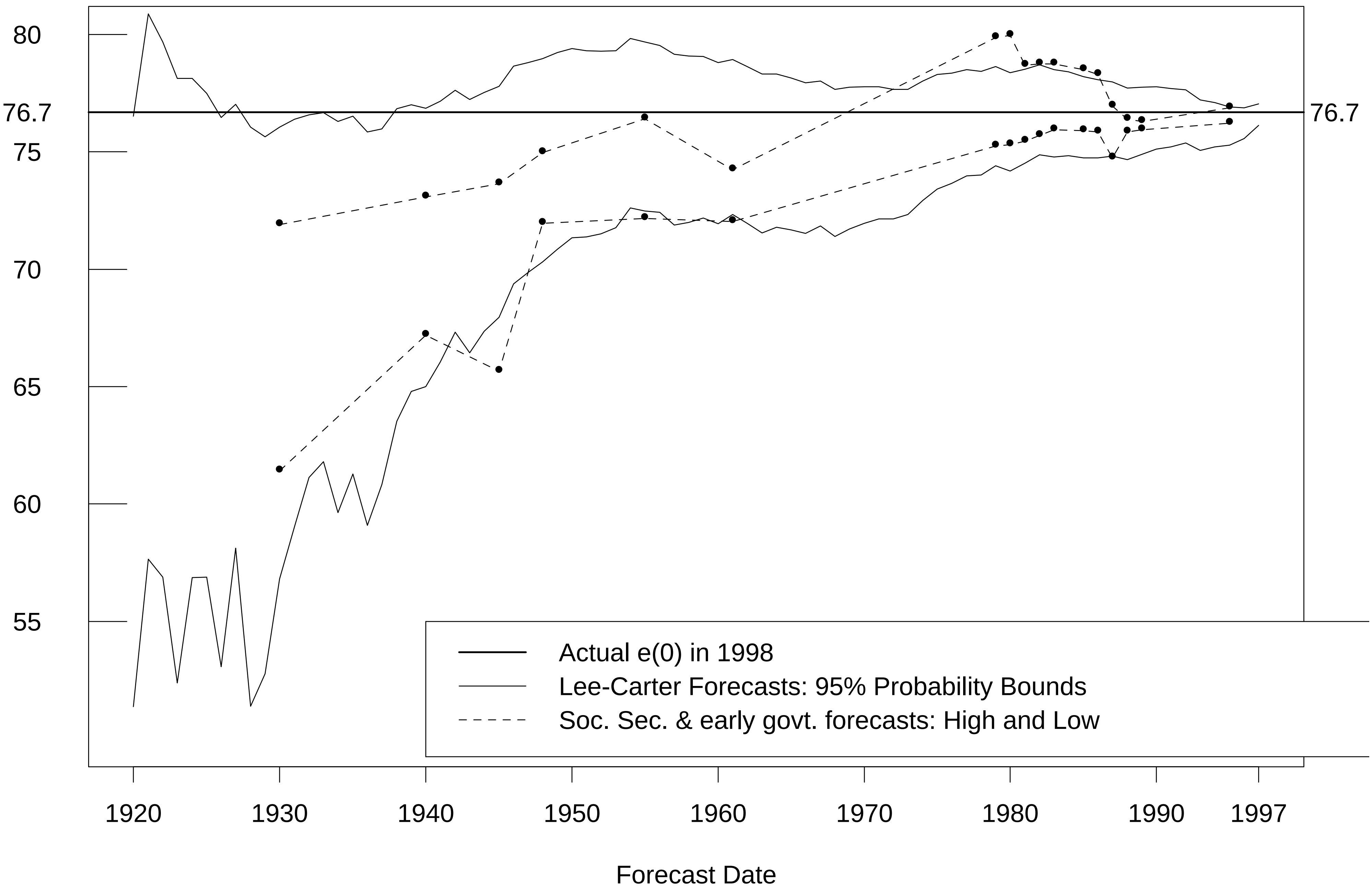

Figure 11: Mean Error in Forecasts of Life Expectancy

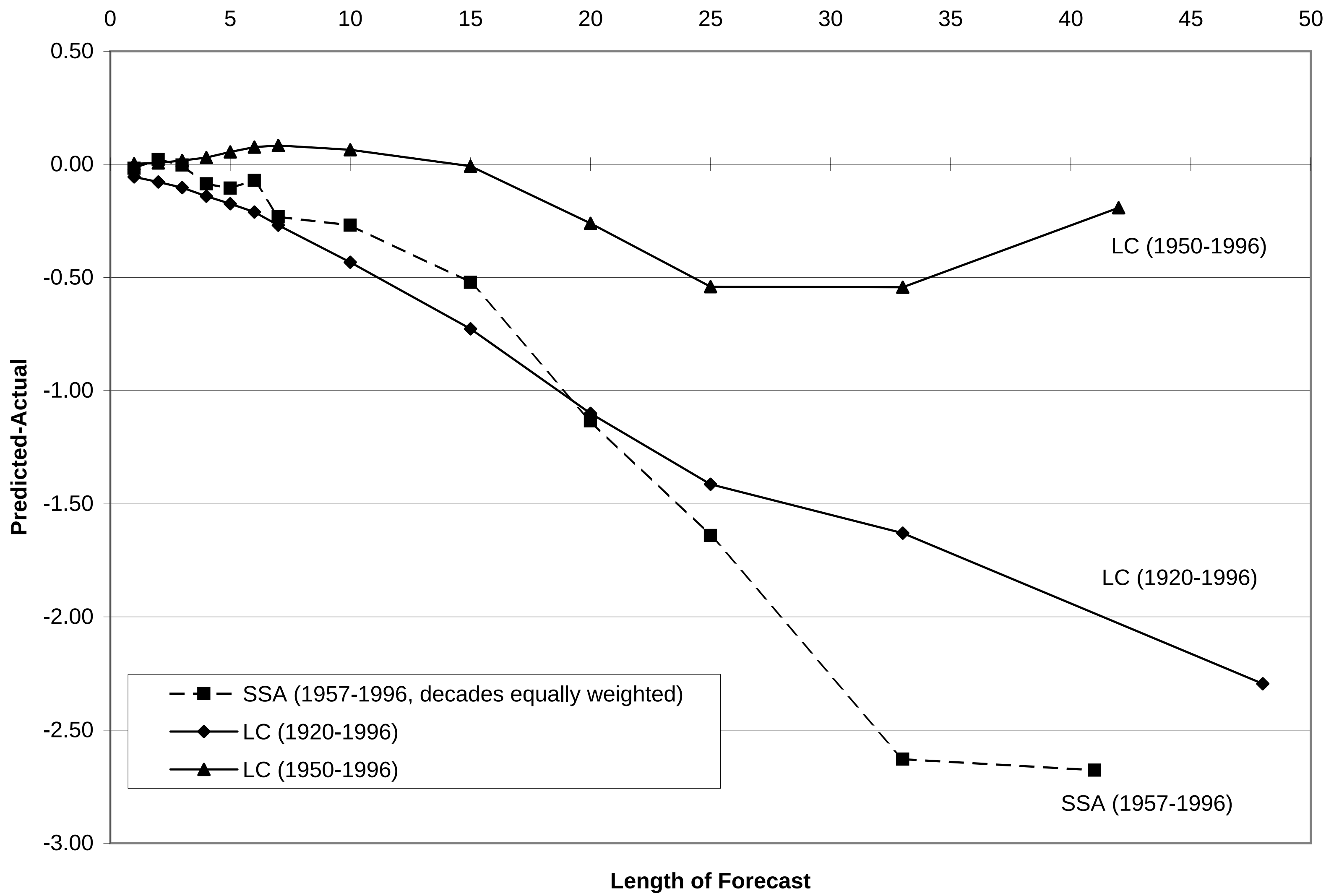

Figure 12: Root Mean Squared Error in Forecasts of Life Expectancy