Title
Population Forecasting for Fiscal Planning: Issues and Innovations

Permalink
https://escholarship.org/uc/item/7n02r268

Authors
Lee, Ronald D.
Tuljapurkar, Shripad

Publication Date
2000

Peer reviewed

December 21, 1998
date last saved: 12/21/98 6:01 PM
date last printed: 03/24/99 2:43 PM

Population Forecasting for Fiscal Planning: Issues and Innovations

Ronald Lee
Demography and Economics
University of California
2232 Piedmont Ave
Berkeley, CA 94720

Shripad Tuljapurkar
Mountain View Research
2251 Grant Road
Mountain View, CA 94024
email: tulja@mvr.org

We are grateful to Michael Anderson, Timothy Miller, Carl Boe, Ryan Edwards, and Bryan Lincoln for their research contributions to projects on which this paper draws. We have benefited from comments on an earlier draft by Dan McFadden, Jim Smith, and Peter Diamond, as well as by other conference participants. Lee's research for this paper was funded by a grant from NIA, AG11761. Tuljapurkar's research for this paper was funded by a grant from NICHD, HD32124. The authors also acknowledge support by Berkeley's NIA-funded Center for the Economics and Demography of Aging.

I. Abstract

This abstract consists of a concise list of conclusions from the analysis in this paper.

A. Assessing Recent Official US Vital Rate Forecasts

- In retrospect, it appears that over the last fifty years, the Census and Social Security forecasters attached too much importance to the most recently observed levels of fertility and mortality.
- Recent Census Bureau projections of US fertility, based on race/ethnic disaggregation, appear to be too high. Recent Social Security fertility projections appear reasonable, although the range may be too narrow in light of international experience.
- Recent projections of life expectancy gains by both Census and Social Security appear to be substantially too low, in light of past US experience and international levels and trends in low mortality countries.

B. Uncertainty in Population Forecasts

- The standard method for dealing with uncertainty in demographic (and many other) forecasts is the use of high, medium and low scenarios. This approach is deeply flawed, because it is based on very strong and implausible assumptions about the correlation of forecast errors over time, and between fertility and mortality. The random scenario method is an improvement, but it retains some of the same flaws.
- Stochastic population forecasts based on time series models of vital rates (Lee-Tuljapurkar) appear to offer some important advantages, although long forecast horizons in demography far exceed the intended use of these models, and it is necessary to impose external constraints on the models in some cases to obtain plausible forecast behavior. One should not rely on mechanical time series forecasts in any case; they should be assessed in relation to external information.
- A parsimonious time series model for mortality appears to perform well within sample in applications in various countries, and suggests future life expectancy gains in the US at roughly twice the rate projected by Census and Social Security.

C. Results of Population Forecasts

- Middle forecasts by Census, Social Security, and Lee-Tuljapurkar (LT) agree closely on the timing and extent of increase in old age dependency ratios as the baby boom ages (although LT are somewhat higher due to lower mortality), but Census shows some amelioration after 2040, due to higher fertility. After 2040, LT forecasts continue to increase, doubling by 2070 to .45, while Social Security forecasts increase to .41. The Social Security range is three times as wide as that of Census, reflecting inherent flaws in the scenario method.

- Middle forecasts of the Total Dependency Ratio by Census and LT agree fairly closely , but are somewhat higher than Social Security (LT is .88 in 2070; Social Security is .83). The Social Security range is extremely narrow, reflecting inherent flaws in the scenario method, but the Census range is far too narrow as well.
- In Social Security forecasts, the correlation between errors in forecasting Youth Dependency Ratios and Old Age Dependency Ratios is close to –1.0. In Census forecasts, it is moderately positive. These correlations result from the bundling of assumptions in scenarios. LT forecasts show a correlation of -.6 to -.4, indicating partially offsetting variations in the proportions of children and elderly, as one would expect.

D. Stochastic fiscal projections:

- We analyze the performance of Social Security projections of cost rates since 1950, for forecast horizons of up to 35 years. Performance was generally very good, with no systematic bias, small average errors, and root mean squared errors smaller than the published high-low ranges. Projections done from the mid-70s to the mid ‘80s have under-projected costs by 12%, however.
- Middle LT forecasts suggest that government expenditures on the elderly will increase by over 150% in relation to GDP by 2070, while expenditures on children and age neutral expenditures will remain flat. Taxes rise from 24% now to 38% of GDP in the median forecast to 2070 (if debt/GDP is constrained), while the 95% probability range for taxes in 2070 goes from 25% to 53% of GDP.
- Increased costs of OASDI account for nearly 30% of the increase in expenditures on the elderly, but a larger share, 57%, is due to health costs in the median forecast. Fixing Social Security will not take care of the long term budget problem.
- Investing 90% of the Social Security reserve fund in equities yielding 7% (real) would fix the system according to a deterministic simulation, but in a stochastic forecast there is still a two thirds chance of exhaustion, with a median exhaustion date of 2044, and a negative median (but strongly positive mean) Fund balance in 2050.
- Raising the payroll tax rate by 2% immediately should nearly put the system in long term actuarial balance according to Social Security projections, but still leaves a 75% chance of fund exhaustion before 2070 in LT stochastic forecasts.
- Raising the normal retirement age to 71 by 2023 raises the median long term actuarial balance above 0 in LT stochastic forecasts, but still leaves a 43% chance of fund exhaustion before 2070.

II. Introduction

Is population forecasting different than other kinds of forecasting, that it should warrant its own special methods, and its own special discussion? In some important respects it is, and in particular, long term demographic forecasts many decades into the future may contain more useful information than is true for other forecasts, such as turning points. There are several reasons. 1) The initial age distribution of the population provides early information about future population size, age distribution, and growth rates. For example, since their birth, we have known exactly when the baby boom generations would swell the numbers of elderly. 2) The relative slowness, smoothness and regularity of change in fertility and mortality facilitate long term forecasts. Compared to real productivity growth or to real interest rates, for example, the vital rates are less volatile. 3) Fertility, mortality and nuptiality have highly distinctive age patterns which have persisted over the several centuries for which they have been observed. These regular and distinctive age patterns reinforce the preceding two points, by making the consequences of initial age distributional irregularities more predictable. Demographers have developed methods and models for exploiting these features of population evolution in their projections. This does not mean, of course, that demographers have built a sterling record of success in long term forecasting. Their record, which we will review later, has been a mixture of success and failure.

Demographic forecasts have many uses. A few users, such as the manufacturers of infant formula, are interested in the numbers of births by quarter in the coming year. Educational planners are interested in the numbers of school age children, typically in a local area, over a longer horizon, perhaps five to 20 years. Some users, such as planners for Social Security and Medicare, have much longer horizon of 75 years, and are particularly interested in the age distributions of workers and the elderly. Social Security planners also need information on the distribution of the future population by marital status, since benefit payments differ by marital status and by living arrangements. Environmental analysts also have long horizons, but are typically less interested in the details of age distributions. This paper will focus on long run forecasts of national populations, and specifically will consider forecasts over a 75 year horizon with detail on age distribution. Sometimes population projections are used for analytic purposes, to consider the effects of different future scenarios, rather than as predictions. Here we will restrict our attention to predictions or forecasts. We believe that most, though not all, population projections fall into this category, despite any disclaimers by their authors.

We will also focus primarily on what might be called core demographic forecasts, of fertility, mortality, migration, population size, and population age distribution. Many other demographic variables are of interest, but discussing them would take us far afield, and dilute our effort. Thus we will not discuss forecasts of marriage, divorce and the corresponding statuses of the population, household living arrangements, and kinship ties. We refer readers to Mason (1996), Goldstein (1997), Wachter (1997) and the Office of the Actuary of the Social Security Administration (henceforth OASSA) (1997) for work and literature review on these topics. Nor will we consider projections of the health, functional status, disability, or cause of death of members of the population. For these we refer readers to Manton, Corder and Stallard (1997), Wilmoth (1996), and OASSA (1992). Forecasts of labor force participation, income, education, and related variables are even further outside our scope.

III. How Demographers Approach Forecasting

Demographers typically approach forecasting through disaggregation. Faced with apparently varying demographic rates, the demographer's instinct is to break the population down into skillfully chosen categories, each with its own corresponding rate. The hope is that by so doing, it will be found that these more disaggregated rates will be found to be constant, or to be varying in regular and predictable ways. If the population growth rate is varying, perhaps the variation results from constant age specific birth and death rates applied to a distorted population age structure. If age specific death rates are varying, perhaps the variation is tamed by looking at these by cause of death. If age specific birth rates are varying, perhaps these are tamed by looking at birth rates by age, parity (number of children already born), and length of birth interval, all broken down by race/ethnic category, for example. To take an interesting specific example, the extremely low fertility in Western Europe might be due to continuing postponements of childbearing rather than a change in the more fundamental ultimate number of births per woman (Bongaarts and Feeney, 1998).[1] This change in timing might be revealed by a disaggregation of fertility by parity (number of prior births) and age. The currently low fertility would then reflect an atypical structure of parity by age in Europe. This strategy of proceeding by disaggregation can be illuminating. However, it is limited by its inability to cope with genuine change in the underlying rates. It is through such genuine change in underlying rates that the population compositions and structures became distorted in the first place, and such changes can be expected to continue in the future.

Certain kinds of disaggregation inevitably raise the projected totals relative to more aggregated projections. This happens because any subgroups of the population which have growth rates above the initial average will grow relative to the other subgroups, and so will receive a greater weight in the average of future growth rates, leading to an increase in the projected average growth rates. The level of the population projections and fertility forecasts of the US Census Bureau rose substantially when it began to disaggregate the forecasts by race/ethnic categories a few years ago (although there other causes as well). Disaggregation of mortality by cause of death has a similar effect, when death rates by cause are extrapolated at their historical exponential rates. The most slowly declining cause-specific death rate, or the most rapidly rising one, then comes to dominate the total death rate in the long run, so mortality is projected to decline more slowly than is the case without disaggregation (Wilmoth, 1995). Pointing out that this is a necessary consequence of certain kinds of disaggregation does not necessarily help us understand whether the higher or lower projection is more correct.

A. Demographic Approaches to Predicting Future Change in Fertility

Economic theories of fertility are highly developed and various models have been estimated and tested. In our view, however, they do not yet provide a useful basis for forecasting fertility. In any event, in order to use any of them, we would first have to develop forecasts of men's and women's potential real wages and non-labor income, and of interest rates, and some key prices, at a minimum.

Nonetheless, there are some basic theoretical (or common sensical) ideas which do influence fertility forecasts. The first of these is the idea that fertility is a means to achieve some desired number of surviving children, at least after the demographic transition is under way. Therefore, declining mortality or reductions in the perceived level of mortality, are expected to cause a corresponding reduction in fertility. Secondly, avoiding births is costly, either in terms of foregone sexual relations or in terms of the steps needed to avoid conception or to abort a conceptus. Consequently, some portion of actual births to the population is unwanted, such that if avoiding births were costless and perfectly efficient, these births would not have occurred. (Correctly accounting for the effects of mistimed pregnancies is a complicated separate issue.) If technological progress brings us closer to costless and perfectly effective contraception, we would expect a decline in the flow of births and in children ultimately born to the average woman. With these two simple and uncontroversial ideas we have reason to expect a long-term downward trend in fertility, without applying more interesting but also more questionable behavioral theories of fertility. Of course, both of these effects have a natural limit which has already nearly been reached in the case of mortality (about 98.5% of births survive to age 20 in the US under current mortality). Unwanted birth rates have also declined greatly in the past 40 years.

How about forecasting change in the desired number of surviving births (completed family size)? One approach is simply to ask women, through surveys, how many children they expect to have ultimately, and when they expect to have them.[2] Analysis has shown that responses are not highly predictive for individuals, but do much better when averaged for age groups. Because childbearing mostly takes place fairly early in a woman's adult life, and because plans change as years pass, data from such surveys are not very informative about fertility more than a few years in the future. Furthermore, if fertility closely follows these plans and expectations, then observing current fertility may provide the same information as the surveys. However, when timing patterns are changing, leading to distortions in the current fertility rate, survey data of this sort may give a truer indication of long run tendencies. Thus survey data for European populations typically show that women want around two children, although the

European Total Fertility Rate (TFR) currently averages only 1.4 children per woman (Bongaarts, 1998).

B. Demographic Approaches to Predicting Future Change in Mortality

There are also behavioral, biological, evolutionary, mechanical and statistical models of functional status and survival (see Manton, Stallard and Tolley, 1991; Lee and Skinner, 1996, Wachter and Finch, 1997; Wilmoth, 1998, and Tuljapurkar and Boe, 1998). With a few exceptions, none is currently a useful basis for forecasting, although they influence the general range of possibilities that must be entertained as possible. The work of Manton and colleagues estimates nonlinear models relating risk factors and life style behaviors to mortality and functional status, such that mortality forecasts can be derived if forecasts of the driving forces are available. In our view, the advantage of this approach lies in its ability to link functional status, disease states, and cause of death in a dynamic structural model, and to use this model to analyze the consequences of certain kinds of policy relevant changes. We do not believe it will provide more accurate long term forecasts of mortality, due to the complexity of the approach, the shortness of the available time series which must be used to forecast life style behaviors, the large number of parameters that must be estimated, and the non-linear way that parameters and forecasted life style behaviors or risk factors interact to generate the mortality forecast.

There are also many empirical studies of mortality change over time, and these make a very useful contribution to the forecasting problem by revealing the pace and pattern of change in death rates by age and sex. For example, it is useful to know that although US female old age mortality has been stagnating for the past fifteen years, elsewhere in the industrial world it has continued to decline rapidly or even accelerated (Kannisto et al, 1994; Horiuchi and Wilmoth, 1995), so there is good reason to expect the mortality decline for older US women to accelerate in the future. The recent stagnation is not a consequence of approaching an upper limit.

C. Historical and International Analogy

Demographic transition theory is a combination of empirical generalization based on the earlier experience of countries that have already achieved low fertility (until recently, largely European), and some generalizations about the influence of socio-economic change on fertility levels. Suffice it to say that this theory is of no use for predicting the future fertility of industrial nations. Some projection procedures for countries earlier in the transition have used curves fitted to the fertility and mortality trajectories of countries farther along in the transition, or that have completed it. These procedures have been surprisingly successful, but are not useful for post-transitional populations.

D. Implicit Assumptions

Population projections are based on a set of assumptions that are only occasionally stated explicitly. Projections assume there will be no catastrophic event such as nuclear war, or a collision with a large comet. They usually also make no provision for more predictable changes, such as global warming. More generally, projections assume that there will be no deep structural change, in the sense that they extrapolate history and expect the future to be like the past in certain respects.

Most projections assume that vital rates vary independently of the distribution of the population across the categories to which they apply. Put differently, it is usually assumed that there is no kind of feedback in the demographic system. Such an assumption rules out the theory advanced by Richard Easterlin (1968, 1978) that larger generations tend to experience economic and social adversity, leading them to have lower fertility, and perhaps causing them to produce fewer births than would a smaller generation. Conventional methods would have predicted a baby bust in the 1950s and early 1960s instead of the actual baby boom, because the parental generations born in the 1930s were small. Similarly, conventional methods would have predicted more births in the late 1960s and 1970s as the baby boom children began to reproduce, rather than the actual baby bust. Easterlin did in fact predict the baby bust, but he also predicted a new baby boom in the later 1980s and 1990s, which never materialized. The dynamic behavior and forecast methods derived from populations subject to Easterlin-type feedback have been studied (Lee, 1974; Lee, 1976; Wachter, 1991, for example). The US Bureau of the Census actually incorporated feedback in experimental population forecasts published in 1975. On balance, although the feedback models are very interesting, there is not sufficient empirical evidence to justify using them for practical forecasts.

Those who believe that the world population is already unsustainably large argue that the environment will bite back in response to further population growth, leading to higher mortality and lower fertility. Sanderson (1995) has modeled and discussed the projection issues raised by this view. Others suggest that if fertility continues for much longer at below replacement levels (as in Europe or Canada) there will be a public policy response in the form of powerful pronatalist policies. Econometric and demographic studies suggest, however, that the ability of governments to affect fertility in industrial nations is quite weak. Romania achieved spectacular increases in fertility when it suddenly outlawed abortion and contraception, but these gains were short-lived, as fertility quickly returned to its earlier levels. Sweden for a time appeared to have substantially raised its fertility through a combination of policies making it easier for parents to rear children without financial or career sacrifice. However, these policies turn out to have affected only the timing of births, and fertility has now fallen back to its earlier levels.

It is, perhaps, surprising that projections of mortality take no account of forecasts of public expenditure on health care or on medical research, even when both are discussed together (as in Lee and Skinner, 1996).

While many of these assumptions may seem extreme, it is really not clear how one could proceed without making them. Generally, we think it reasonable to proceed in this way.

IV. Assessing Performance of Past Projections

A. Census Projections of US Fertility

Traditionally, the Bureau of the Census has focused its best efforts on the fertility forecasts, while the Office of the Actuary of SSA has focused on the mortality forecast. For this reason, we will examine the past record of BC for fertility projections, and of OASSA for mortality projections.

[Figure 1 here].

Figure 1 plots all the forecasts made by the US Bureau of the Census since 1949. Where a middle forecast was given, we have plotted that. Where no middle forecast was given, we plot the average of the two middle range forecasts. We also plot the actual TFR for each year over this period. The methods used to make these forecasts have relied to varying degrees on extrapolation, professional judgement, survey data on birth expectations, and on basic insights from sociological and economic theory. In our view, the fertility forecasts correspond not only to the view of this official agency, but also reflect the prevailing opinions of professional demographers. The forecaster’s are competent, and we do not mean to suggest that any of these forecasts was a bad guess in its historical context. The figure shows clearly the severe limits on demographers’ ability to forecast fertility. Every turning point is missed, and by and large, the projections simply mimic the level of fertility in the years immediately preceding the forecast. Indeed, the ultimate level of the fertility forecast is correlated +.96 with the average TFR in the five years preceding the published forecast! It is particularly striking that the forecasters do not have in mind a central value towards which the forecasts converge over time. The ultimate forecast levels range from 1.8 to 3.4 births per woman.

Recent fertility forecasts by BC foresee an ultimate TFR of 2.245 (Day, 1996:2, Middle Series). In our view, this forecast is unrealistically high. It follows from the assumption that there will be no change or convergence in the fertility of any race/ethnic group between 1995 and 2050 (Day, 1996:2). The projected increase in the TFR from the current 2.055 to the future 2.245 is due entirely to projected changes in the race/ethnic composition of the population. However, research has shown that when fertility is examined by immigrant generation, there is strong convergence to the level of non-Latino whites after two generations (Smith and Edmonston, 1997). The persistence of high fertility of immigrant groups will therefore depend on first and second generation immigrants remaining a constant share of the total membership of the Asian and Latino race/ethnic groups, which is consistent with forecasts (Smith and Edmonston, 1997). However, as fertility drops in countries sending immigrants to the US, which has been occurring in recent decades, we would expect the fertility of entering immigrant women to have lower fertility on arrival. Fertility in East Asia has already dropped below replacement in many countries. The Mexican TFR has dropped from a high near 7 in the late 1960s to a current level below 3, with the UN predicting it will reach replacement level around 2015. Yet BC is projecting that the TFR for Hispanic women will remain at 2.98 births per woman until 2050 (Day, 1996:A7). These actual and predicted changes seriously undermine the BC projection of constant fertility within race/ethnic groups. The Low assumption for fertility by BC also appears unrealistically high at 1.91 (Day, 1996:4), in light of the much lower fertility in Europe, and lower fertility in the US in the 1970s and 1980s. OASSA assumes (Intermediate) a TFR of 1.9, which we believe is reasonable, with a range of 1.6 to 2.2.

[Figure 2 here].

Figure 2 is based on the same set of projections, but it shows the high and low brackets for each forecast, and does not show the middle. Eleven brackets are shown. For five of these eleven brackets, actual fertility has escaped the high-low bounds *within three years of the base year*(!); in at least one case (1972), this was before the projection was even published. It is not generally stated what the probability coverage of these brackets is intended to be, but presumably the authors would regard these brackets as having failed, since more than half were wrong within three years.

But what is the intent of brackets of this sort? Because they are used to define the high-low range for long run brackets on population size and other variables, one might argue that they are intended to bracket the long run averages for fertility, but not necessarily to capture all year to year fluctuations. On this view, one could not say they had been unsuccessful until many decades had passed. However, the violations of bounds in Figure 2 are not typically the result of minor blips, but rather seem to reflect longer run changes. Have forecasters learned from this record? It appears that forecasters quickly forgot about the past volatility of fertility, and were lulled by the period of stability between 1975 and 1987, narrowing their brackets as the baby boom faded into the past. The bracket for the 1985 forecast was violated within a single year.

Some indication of the uncertainty about future fertility may be drawn from analysis of the historical record, including the low fertility of the 1930s, the high fertility of the baby boom, and the low fertility of the baby bust. This records suggests that the small variation of the past two decades should not lull forecasters into complacency. Until we understand the causes of the baby boom we should not dismiss it as a one-time anomaly. Comparison with other industrial and industrializing countries also indicates that caution is called for. The average TFR in Europe is only 1.4 births per woman, and for all the “developed” country populations including the US, it is still only 1.6. Some countries have TFRs around 1.2 or 1.3 (Spain, Italy, Germany, Hong Kong). It is still too early to say whether these low levels of fertility primarily reflect timing distortions of the sort discussed earlier, or whether they indicate a long term low level or even a continuing trend toward still lower fertility. Under these circumstances, it would be prudent to consider the possibility that US fertility may be lower in the long run than the 1.6 children per woman assumed in the OASSA "high cost" projection.

B. Social Security Projections of US Mortality

[Figure 3 here].

Figure 3 plots the average of male and female life expectancy projections (intermediate, when more than one is available) done by the OASSA from 1945 to the present. Forecasts made between 1945 and 1965 were quite accurate until the early 1970's, when mortality began to drop more rapidly, and life expectancy to rise more rapidly, than anticipated. The period of rapidly rising life expectancy left all the earlier forecasts in error by two or three years, a discrepancy that has persisted up to the present for those earlier projections. Not surprisingly, projections made just before or just at turning points in the rate of change of mortality have fared the worst. Thus the 1974 projection most thoroughly reflects the belief that the slow gains from 1955 to 1968 would continue into the future. The projection done in 1983, just at the end of the period of rapid mortality decline, most thoroughly reflects the belief that the period of rapid gains would continue, leading to early errors of about one year in e0. Examination of the separate forecasts for males and females reveals similar patterns, but with larger errors for females than for males.

This review of the past record of OASSA does not suggest any systematic tendency to project life expectancy gains that are too large or too small. Yet in our view, in recent years OASSA has been predicting gains that are too low, and also has predicted a peculiar age distribution for mortality decline. Here we will discuss these points briefly.

[Table 1 here.]

Table 1 compares the average rate of decline of death rates for broad age groups of the population to the rate of decline that OASSA forecasts for these groups over the next 85 years. We see that overall, the rate of decline that is projected is less than half as rapid as that observed in the past (.57% versus 1.18% per year). When we look at the age pattern of the discrepancies, we see that they are greatest at the younger ages, and decline to near zero for the 85+ category. The overall death rate would be 67% higher in 2080 under the OASSA projection than under trend extrapolation,. The death rate for children would be 4.4 times as high, for working ages would be 1.8 times as high, for the younger elderly would be 1.4 times as high, and for the oldest old would be nearly the same. It is not clear why this change in either the pace of mortality decline, or in its age composition, is projected.

Comparison with international mortality trends also suggests that the OASSA projections of life expectancy are too low. The population of Japan currently has a life expectancy of 80.3 years. According to OASSA projections in recent years, the US will not reach 80.3 until just before 2050 (for example, see Trustees, 1998:60), which seems unduly pessimistic. According to the BC Middle Series, life expectancy will reach 80.25 in the US in 2035. This also seems to us to be too pessimistic. A careful study of mortality trends at ages 80 to 100 in 19 countries with reliable data concludes that "In most developed countries outside of Eastern Europe, average death rates at ages above 80 have declined at a rate of 1 to 2% per year for females and 0.5 to 1.5% per year for males since the 1960s." (Kannisto et al 1994:794). OASSA, however, projects a future rate of decline at ages above 85 of only .5% per year (see Table 1), which is less than half the average pace in the Kannisto et al (1994) populations. Kannisto et al report that the rates of mortality decline at these high ages have been accelerating throughout the century. There is also little evidence that populations with lower mortality at these advanced ages are experiencing less rapid declines.

A study by Horiuchi and Wilmoth (1995) of a smaller set of industrial nations reaches similar conclusions for mortality at ages 60 to 80 over recent decades. The combination of historical trends within the US, and international trends outside the US, provides compelling reason to believe that the OASSA life expectancy projections are too low.

It is important to note, however, the particular age pattern of mortality decline projected by OASSA (see Table 1). Relative to trend, they project the slowest declines at younger ages, whereas the discrepancies at higher ages are smaller or nonexistent. If one were to switch to a trend extrapolation forecast (see the discussion of Lee and Carter, 1992, below), survival through the working years would be substantially higher than in OASSA, and this would partially offset the increase in overall life expectancy.

V. Dealing With Uncertainty

Long term demographic forecasts are obviously highly uncertain, as are most other kinds of long term forecast.

A. Scenarios

The most common means of assessing and communicating uncertainty, in demographic forecasting as in other kinds of forecasting, is to formulate high and low trajectories for the key inputs to the forecast, to combine these into collections of input trajectories called "scenarios", and then to prepare and present the results of at least two such scenarios in addition to the preferred forecast. Often these alternate scenarios are identified as "high" and "low" in some sense. As examples of this procedure, the BC bundles high fertility, low mortality, and high net immigration into a "high" scenario, because all the trajectories conduce to a high future population size or growth rate. OASSA, by contrast, bundles low fertility, low mortality, and low immigration into a "high cost" scenario, because these choices all conduce to a higher old age dependency ratio and higher costs per tax payer for the system.

The scenario approach does not attach any probability coverage to the forecast bands, and for good reason. Any probabilistic interpretation of the scenarios would founder immediately on inconsistencies. To provide probability bounds for fertility or births in each year, the brackets would have to be wide enough to contain annual blips and drops; but most of this high frequency variation would cancel out and be irrelevant for the longer run evolution of the population. This kind of problem infects the brackets for almost all demographic variables that are forecast. Age groups involve summing over births and deaths in individual years, so brackets should be proportionally smaller. Births result from applying birth rates across a broad range of age groups, and again there should be averaging of errors and brackets should be proportionately smaller.

[Table 2 here].

A related kind of problem comes from the need to bundle alternative trajectories into scenarios, with choices made about how to bundle them, as illustrated by the description of BC and OASSA procedures earlier. Table 2 shows the range of uncertainty for BC and OASSA projections published in 1992, with a time horizon of 2050. The numbers in the table are the difference between the high and the low projection, divided by twice the middle projection, expressed in percents. The BC column indicates a high degree of uncertainty for the number of children, the number in the working ages, and the number of elderly. However, near certainty is indicated for the old age dependency ratio (OADR), because high fertility leads to more workers, and low mortality leads to more elderly, so bundling the two together in a scenario yields very little variation in the OADR. The Total Dependency Ratio (TDR) has larger variation, because it is additionally affected by variations in the number of children, which are not offset. However, even for the TDR, the range appears to be inappropriately small, given the uncertainty about its constituent parts.

The OASSA column shows less uncertainty about each of the population elements forecast, but the OADR has a range seven times as great as that of the BC. The reason is clear: OASSA bundles low fertility with low mortality, and so the uncertainties in the two reinforce, rather than offset, one another. But now the TDR has uncertainty near zero! These kinds of internal inconsistencies are an intrinsic feature of attempts to represent uncertainty through scenarios. The problem is that BC assumes a perfect negative correlation of errors in forecasting fertility and mortality ($\rho=-1.0$) while OASSA assumes a perfect positive correlation ($\rho=+1.0$). In truth, there is little basis for assuming any correlation between the two at all. The last column shows the 95% intervals from stochastic population forecasts to be discussed later, in which the correlation is assumed to be zero.

B. Random Scenarios Based on Expert Opinion

A new approach developed by Lutz, Sanderson and Scherbov (1996, 1997) and Lutz and Scherbov (1998) seeks to avoid these difficulties through a “random scenario” approach. In this approach, the high, medium and low trajectories for fertility, mortality, and migration are mixed by randomly choosing trajectories independently within the high-low range. The trajectories for each variable maintain their shape across time, but are multiplied up or down by $(1+\varepsilon_{i,j})$ where i varies over fertility, mortality and migration, and j indicates the particular random scenario to be simulated. Note that $\varepsilon_{i,j}$ does not vary with t, that is over the forecast horizon. Fertility, for example, will always be somewhat high, or somewhat low, over any particular random scenario. Through random simulation, a set of many random scenarios is generated. Then the appropriate summary statistics (mean, median, probability distributions) can be calculated from this set. Note that the initial high, medium and low trajectories for the rates are taken as given by this method. In actual applications they have either been developed through consultations with panels of experts, or they have been taken from the ranges provided by government statistical agencies responsible for preparing projections, counterpart to our BC or OASSA.

This approach does indeed seem preferable to the traditional scenario approach, since it avoids the false assumption that fertility and mortality forecasting errors are correlated either +1.0 or –1.0. However, it still assumes that errors in fertility (and mortality) are perfectly correlated over time. If fertility is higher than expected in the first few years of some random scenario, it will be higher than expected for all future years ($\varepsilon_{i,j}$ does not vary with t). In real life, however, fertility rises and falls in unpredictable ways, and mortality declines sometimes rapidly and sometimes slowly. Random scenario forecast sets generated in this way will never allow for the possibility of a baby boom or a baby bust, as episodes occurring along an other wise medium trajectory. They cannot possibly represent correctly the variance-covariance structure of population forecasts. Whether their deficiencies lead to important quantitative distortions, or to negligible ones, has not yet been established.

Pflaumer (1988) proposed a different approach, which avoids assuming perfect correlation of vital rates in a give year, and also avoids assuming perfect intertemporal correlation of errors in each vital rate. He used Monte Carlo methods to draw random values for each vital rate in each time period, assuming some probability distribution for the vital rates within the high-low range that was taken from official forecasts. Pflaumer assumed there was no autocorrelation of forecast errors in the vital rates, which is inconsistent with very high empirical estimates of autocorrelation in fertility, and in rates of change in mortality. With zero autocorrelation, most of the year-to-year variance in the vital rates averages out over time, and consequently probability bands from this method appear to be far too narrow.

These difficulties in converting expert views on middle trajectories and high-low ranges into probabilistic projections raise troubling questions about the expert opinions themselves. What question does the expert try to answer, when asked for a 90% probability range for fertility in 2030? Does the expert seek a range which will contain 90% of annual values for 2030, or which will contain 90% of the long run average fertility trajectories? This apparently minor distinction alone makes a difference of 40 or 50% in the width of the interval, based on a fitted statistical model for US fertility (Lee, 1993). Does the expert have in mind an autocorrelation structure for the errors? Aside from these questions of interpretation, one might wonder whether an expert would be capable of sensibly guessing at probability bounds with coverage of 90% versus 95% or 99%. We would have great difficulty doing this ourselves.

C. Analysis of Ex Post Errors

Another approach is to use ex post evaluations of the sort produced by Keyfitz (1981) and Stoto (1983) to develop probability bounds for the growth rate and size of the projected population. Stoto concluded that an optimistic standard error for the for the annual growth rate forecast by the BC was .3%, based on United Nations projections for developed countries, and a pessimistic standard error (based solely on US BC forecast performance) would be .5%. The BC itself estimates a mean square error for a 10 year horizon of .31%, consistent with Stoto's optimistic interval, and for 20 years of .45%, consistent with Stoto's pessimistic interval (Day, 1996:30). BC does not report standard errors analogous to Stoto's, so direct comparison may be misleading.

VI. The Time Series Approach to Forecasting Vital Rates and Population

A small literature has developed a different treatment of uncertainty in population forecasting, based on the analysis of historical time series of fertility and mortality. This literature is discussed and evaluated in Lee (forthcoming). Besides the present authors and their collaborators, whose work is discussed below, the main contributors have been Alho (1990) and Alho and Spencer (1985, 1990), Cohen (1986), and McNown and Rogers (1989, 1992).

Over the past decade, Lee, Carter and Tuljapurkar have developed the time series approach to population forecasting in a series of articles. Lee and Carter (1992) developed a statistical time series model of mortality and Lee (1993) developed a related model for fertility. These were subsequently used to produce stochastic population forecasts by Lee and Tuljapurkar (1994), which will be discussed at length below. Here, we will briefly discuss the time series models of fertility and mortality.

A. Mortality

Let m(x,t) be the death rate for age x in year t. Let a(x) and b(x) be age specific but time invariant parameters, and let k(t) be a parameter that varies over time but is independent of age. The model used by Lee and Carter was:

$$\ln(m(x,t)) = a(x) + b(x)k(t) + \varepsilon(x,t).$$

None of the variables on the right is directly observable, but the model has a least squares solution which can be found, for example, by using elements of the singular value decomposition (SVD).[3] The model in fact has given a very good fit for the time series of age specific death rates to which it has been applied. For example, it accounts for 97.5% of the variance over time in the age specific death rates in the US, 1933 to 1987, excluding the rate for the open interval, 85+.[4] Gomez de Leon (1990) selected this same model in an independent exploratory data analysis of the long historical Norwegian mortality data set. It is important to model the log of the death rates, because otherwise projection leads to negative death rates.

When k(t) declines linearly, each m(x,t) declines at its own exponential rate, b(x)dk/dt. The strategy is to model the time series k(t) using standard statistical time series methods. When this is done for the US, a random walk with drift works quite well, and this is also true for some other countries. The fitted model can then be used to forecast k(t).

[Figure 4 here].

Figure 4 plots the fitted values for k(t) for the US, 1900 to 1996[5]. The basic linearity of the decline in k is striking, despite some fluctuations about the downward trend. Surprisingly, the decline in k in the first half of the period is almost exactly equal to the decline in the second half of the period. By contrast, life expectancy declined twice as much in the first part of the period as in the second, but lives saved by falling mortality shift increasingly to older ages, where the increment to life expectancy is smaller because there are fewer remaining years in any case. The figure also shows the 95% probability bounds for the forecast of k. The uncertainty in the forecast of k includes three components: the innovation term in the fitted random walk process of k; uncertainty of the estimated rate of drift in the random walk process; and a 1/97 chance each year that an epidemic similar to the flu epidemic of 1918 (a 6 unit increase in k) will occur.[6] In future work, we plan to take into account the fitting errors in the basic model, as well. Using the equation given above, the death rates for each age in each future year can be calculated, and from them, any desired life table functions can be found.

[Figure 5 here].

Figure 5 plots the resulting forecast for life expectancy for each year, with its 95% interval. Life expectancy is forecast to rise roughly twice as fast as under the OASSA projection, which has it rising only to 81.5 by 2070, versus 86.0 here, from a current level of just over 76. The high OASSA forecast is slightly below the mean forecast for Lee-Carter. Recall, however, that the Lee-Carter age pattern of decline projects higher survival in the working years relative to OASSA, so that implications for Social Security finances are less severe than one might expect.

[Figure 6 here].

Figure 6 compares the Lee-Carter (1992) and OASSA (1992) forecasts of life expectancy at birth to the actual, for 1990 through 1997; both use data through 1989.[7] The OASSA forecasts are indicated by triangles, and can be seen to be on the low side in most years. The Lee-Carter published forecasts are marked by diamonds, and they are all too high. There is, however, an important difference in the source of these errors in Lee-Carter and OASSA. OASSA errors result from a rate of increase that is lower than the actual. The Lee-Carter errors result from an error they made in estimating life expectancy in the jump off years 1988 and 1989, for which they did not have access to the actual age specific mortality data, and instead inferred mortality in those years indirectly from the published numbers of deaths. This led to an overestimate of life expectancy for 1989 by about .3 years, and this error persists in their forecasts. Although wrong in their baseline data, Lee-Carter do appear to have gotten the rate of increase correct. The dashed line, and its bounds, is the forecast that Lee and Carter should have made had their baseline data been accurate.[8]

We should note the following features of this approach to forecasting mortality:

- Statistical time series models in the Box-Jenkins tradition were never intended for long term forecasting. They rely on simple, low order linear approximations to processes that may be much more complicated. The approximations may work well for forecasting a few periods ahead, but there is no good reason to expect them to perform well far into the future.
- Our model implicitly assumes that variations over time in age specific rates are highly correlated across age. This is, in fact, quite a good assumption.
- Forecasting errors for mortality will arise from errors in forecasting k(t), and these in turn depend on the explicit innovation term for the random walk process, errors in estimating the drift term, and any errors of specification and conceptualization. Errors from k(t) are likely to dominate errors from $\varepsilon(x,t)$, and from estimation of the a(x) and b(x), after a few decades (see appendix to Lee and Carter, 1992).
- The trend fit to mortality (here described by the drift in k(t)) may depend sensitively on the time period over which k(t) is estimated. For the US, it was most recently estimated from 1900 to 1996. In the 19th century, the rate of mortality decline was much slower, and indeed sometimes mortality rose over fairly long periods. How should we pick the relevant period for fitting? It is, perhaps, deceptive to lean heavily on a period defined by data availability. Fortunately, it makes little difference to the US forecast which start date after 1900 is chosen, until one has moved all the way up to 1960.
- Horiuchi and Wilmoth (1995) find that in recent years death rates at older ages have been declining more rapidly than at younger ages, reversing the earlier pattern. Switches of this sort are inconsistent with the simple model used here, but whether this is a serious problem for the method, or a minor one, is not yet clear.
- McNown and Rogers (1989, 1992) have taken a different approach in a series of articles in which they fit a multi-parameter nonlinear curve successively to each cross sectional set of death rates, and then model and forecast the time series of parameters to generate mortality forecasts. Recent reviews and evaluations by HCFA (Foster, 1997), by a NIA/NAS workshop (Stoto and Durch, 1993) and by Bell (1997) have favored the Lee-Carter approach. Also see Tuljapurkar and Boe, 1998.

B. Fertility

Building on earlier work (Lee, 1974), Lee (1993) developed a time series model for US fertility. It is conceptually similar to the mortality model just discussed, although the birth rates are not logged. Fertility trends are somewhat different at the older and younger ages, and modeling the log of the birth rates gives too much weight to the very low rates for older ages, making the estimated fertility index (analogous to k(t)) behave quite differently than the TFR.

There are two special problems in modeling fertility. First, an ordinary time series analysis leads to long run forecasts of fertility that converge to its sample period mean, at a TFR of 2.65 children per woman. However, the actual TFR has been at or below 2.0 for the past 25 years, and the rest of the industrial world has far lower fertility, so 2.65 seems unrealistically high. Lee (1993) argued that there had been structural change in fertility, due to mortality decline during the 20th century, improved contraception, and perhaps other changes such as rising female educational attainment and male-female wage convergence. Because of these structural changes, it made sense to impose a lower ultimate fertility level on the time series process than had been observed in the past, based on extraneous information. For this reason, constrained mean models were used to fit and forecast US fertility.[9] Lee (1993) took a mean midway between the BC and OASSA Middle assumptions. In recent work, Lee and Tuljapurkar (1998a, 1998b) have taken the long run mean to equal the Middle assumption of OASSA, or 1.9 children per woman. Tuljapurkar and Boe (1997) have examined the sensitivity of the forecasts to variations in the assumed means, as well as to prior and empirical distributions for the level of the constrained mean. In many cases, the results appear relatively insensitive to such variations. They also found in a retrospective validation test that stochastic fertility intervals perform well in the sense of containing realized future fertility, unlike official fertility forecasts.

Standard diagnostic methods indicate that the fertility series should be differenced before modeling. When this is done, probability bounds on the forecasts widen rapidly with forecast horizon to include impossibly high levels of the TFR, and negative ones. Lee (1993) argued that for long term forecasts, it makes sense to use the point estimate of the first order autoregressive coefficient, which here was slightly below unity, rather than first differencing. When this is done, both the forecasts and their probability bounds appear reasonable.

[Figure 7 here].

Figure 7 plots the historical series together with the forecast and the probability bounds, based on a constrained mean of 1.9, with limits of 0 and 4.0, and an ARMA(1,1) process. Note that the width of the 95% probability bounds widens to .8 to 3.0 children per woman by 2040. This 2.2 child range is much wider than the OASSA range of 1.6 to 2.2, or .6 children. The figure also plots the probability bounds for the cumulative average of fertility along each sample path up to each date, which is arguably a better basis of comparison to OASSA's intervals. The 95% bounds for this by 2040 are 1.4 to 2.6, for an interval of only 1.2 children. But these intervals are still not comparable to those for OASSA's scenario method. If OASSA were aiming for a 95% probability coverage by their combined intervals of the eight inputs that vary, that probability coverage would have to apply to the joint variation of the eight, of which fertility is just one. The fertility interval would therefore need to have much less than 95% coverage, and therefore would need to be considerably narrower than its 95% range. Taking these considerations into account, the stochastic fertility range does not seem unreasonably wide. Some may think that a TFR of 3 is out of the question for the future. It is important to keep in mind the message of Figure 1, however: reality often violates our prior notions of what is plausible.

Without doctoring, time series forecasting methods do not perform well for fertility. After doctoring, one might wonder what useful information is provided by the forecast that was not extraneously imposed on it. In our view, the autocovariance structure of the fertility process is the key information that is derived from the time series analysis.

C. Migration

Modeling and forecasting net immigration would encounter many of the same difficulties as for fertility, and similar methods could be used to circumvent them. However, because immigration is so subject to policy decisions, we prefer to take the immigration trajectory as given rather than to forecast it stochastically. Lee and Tuljapurkar (1994) assumed that immigration follows the level and age pattern of the Middle OASSA projection.

D. Comment on Time Series Models for Long Run Vital Rate Forecasts

We do not believe that fitting time series models to indices of fertility and mortality is a panacea for demographic forecasting. Mechanical approaches can produce absurd results, and judgement must enter the process at many points. The long horizons for which we apply time series forecasting methods violate common rules of thumb, such as not to forecast over horizons longer than a third the length of the sample period. The time series forecasts must be assessed in the context of other information from biology, economics, demography and sociology. There may be situations in which they fail to give acceptable results in terms of sample period performance or forecast performance. Certainly they require modification for application to populations that have not completed the fertility transition, in the sense that trend dominates fluctuation during the sample period or is expected to in the future.[10] In the case of mortality, the declining trend could continue for a very long time at the rates projected for the US, so the assumption of a constant rate of drift may not be a problem over the next century.

E. Stochastic Population Forecasting

Given the stochastic models for fertility and mortality just described, the assumption about immigration, and an initial population age distribution, it is straightforward to construct a stochastic population projection, that is a single random sample path. To do this, we simply draw random variables for each year to determine fertility and mortality based on the models, and using the usual demographic accounting identities. This process can be repeated many times – say 1000—to form a set of sample paths for the population and vital rates. From these we can calculate the means, medians, and probability distributions for any quantity of interest.[11]

[Figure 8 here].

Before considering the actual forecasts, it will be helpful to compare these forecasts to those by BC and OASSA, particularly as regards the treatment of uncertainty. Figure 8 is a scatter plot of the level of the TFR in 2050 against the level of life expectancy in the same year. This is shown for the Lee-Tuljapurkar (LT) projections, where x's indicate the 2050 values on each of 1000 sample paths, and the solid diamond indicates the average of these. Solid squares indicate the high, medium and low BC scenarios, and solid triangles indicate the OASSA high cost, intermediate and low cost scenarios. Note the following: a) the 1000 points from LT sample paths form a circular cloud, while the three points representing the scenarios for BC and OASSA each fall on a straight line, reflecting the assumed correlation of +1.0 for BC, and –1.0 for OASSA. b) The LT range for the TFR (vertical distance) is very much greater than the fertility range for either BC or OASSA. LT assigns some small probability to TFR values near 0 and near 4, although TFRs outside that range are assigned probability zero.[12] c) The LT range for e0 is narrower than is BC's, but is wider than that for OASSA. The LT e0 distribution is centered higher than that of BC or OASSA. d) The lines joining the BC and the OASSA scenarios are not orthogonal, because the e0 range for BC is so much wider, which flattens the slope of the connecting line relative to OASSA. e) The random scenario method, described earlier, could be based on OASSA scenarios. Its randomly simulated points would then all lie within a circle with center on the intermediate cost scenario, and radius equal to the distance from that point to the high or low cost scenario (assuming these are symmetrically spaced). Alternatively, the random scenario method could be based on the BC scenarios in the same way.

[Figure 9 here].

Figure 9 is a scatter plot of the TFR in 2050 against the TFR in 2020. Again, this is shown for each LT sample path, as well as for the BC and OASSA scenarios. The LT x's are now contained in a large circle, roughly speaking.[13] As expected, fertility is perfectly correlated at these two dates in the BC and OASSA scenarios, which highlights one of their major weaknesses. The random scenario method would merely fill in the line between the markers for BC or OASSA as the case may be; here there is no circle. When a similar plot is done for e0 in 2050 by e0 in 2020 (not shown), the general features of the chart are the same, but now the LT scatter forms an oblong with a positive slope, twice as long as high. This shape reflects the greater persistence of random shocks in the mortality model, because it is a random walk.

F. Forecasts of Population Aging and Total Dependency Ratios in the US

[Figure 10 here].

Figure 10 plots the LT forecasts of the old age dependency ratio (OADR) with 95% probability intervals, together with the BC and OASSA projections for comparison. Focusing on the central forecasts, we note that all agree about the sharp upswing in the OADR from 2010 to 2030, when the baby boom generations born 1946 to 1965 are turning 65. After 2030, the lower mortality projected by LT leads to higher ratios than the other forecasts. (Note that the BC ratio is defined over a denominator of 18 to 64 year olds, compared to 20 to 64 for OASSA and LT). For OASSA and LT, the OADR continues to rise after 2040, when all the baby boom generations have already turned 65, and have begun to die out. In the BC forecast, the ratio begins to fall again after 2040, due to the higher fertility assumed by BC. It is important to realize that population aging in the US and elsewhere is not a transitory event due to the baby boom, but rather is a permanent and probably continuing change that is punctuated by the baby boom's retirement. The median LT forecast has the OADR rising to .45 in 2072, more than twice the initial level of .21. This exceeds the OASSA forecast of .41.

Now consider the range of the forecasts. Here we note that LT has the widest interval(.29 to .79 in 2072), but that it is not so much wider than that of OASSA. The shocker is the interval for BC, which is less than a third the width of the OASSA interval, and less than a quarter the width of the LT interval. The reason is that BC bundles high fertility and low mortality together, as discussed earlier, in contrast to OASSA. BC is well aware of the problem, and provides a separate scenario to give a more reasonable indication of uncertainty for the OADR, but this discrepancy reflects the deep problems with scenarios.

[Figure 11 here].

Figure 11 shows the total dependency ratio (TDR) in a similar way. (Note again that BC uses a different definition of working ages, which accounts for the lower level.) LT forecasts a TDR rising to .88 by 2070, higher than OASSA's .83. The differences in central forecasts are not large, but the differences in intervals are striking. First, note the much wider intervals for the LT forecast (.66 to 1.18 in 2072), due to the effect of the wide fertility range on the child dependency ratio. Second, note that the OASSA range is extremely narrow, with practically no range at all in 2050. Again, this is due to the bundling strategy. Although the OASSA projections are focused on the old age dependency ratio, they are also used for many other purposes such as general fiscal forecasting and academic analyses. Cutler et al (1990), for example, calculate Support Ratios for OASSA alternative scenarios. These are refined total dependency ratios (inverted), in which the numerator is the earnings-weighted population age distribution, and the denominator is the consumption weighted population age distribution.

It is also worth considering the correlation of variations in forecasts of the youth dependency ratio and the old age dependency ratio. Plotting the OASSA forecasts to 2070 for the YDR against the OADR reveals a nearly perfect negative correlation, resulting from their bundling assumptions. The positive correlation found when this is done to 2050 for BC is not as striking. The correlation for 2070 across the LT stochastic sample paths is -.4. Neither the positive correlation in

BC, nor the correlation near –1 in OASSA, correctly indicates the risk of a generational squeeze in which working generations must support large numbers of both children and the elderly, nor the chances of the opposite outcome. This limits their usefulness for fiscal projections.

G. Are the Probability Intervals Too Broad?

[Table 3 here].

The probability intervals for the LT fertility forecasts are very broad relative to BC and OASSA. Does this make the LT probability intervals for population forecasts inappropriately wide? Let us compare LT standard errors of population growth rates to the ex post root mean squared forecast errors for Census Bureau projections.[14] These can be calculated from Stoto (1983), and are provided by Day (1996:30) which includes more recent projections. Table 3 shows results for horizons up to 20 years. Day's (1996:30) ex post root mean squared errors are twice as wide at each horizon as the LT ex ante standard errors. Stoto's (1983) ex post standard errors are three times as wide as the LT ex ante ones. We can also compare the width of the Lee-Tuljapurkar (1994:1185) ex ante 95% probability interval for population size in 2050 with the ex ante high-low population size forecast of Census (Day, 1993); they match almost exactly. Since the Lee-Tuljapurkar method for generating the median or mean forecast is very different than that of Census, there is no reason why the stochastic ex ante standard errors should match the ex post or ex ante uncertainty of Census forecasts. However, these comparisons do make it clear that the LT probability intervals for population size or growth rate are not wider than either the ex post or the ex ante census intervals, despite the much greater ranges for fertility that are incorporated in the stochastic forecasts. The stochastic intervals are, quite appropriately, wider for other demographic quantities such as the Total Dependency Ratio that involve less averaging and cancellation across errors than do population size or growth rate.

VII. Stochastic Demographic Forecasts Applied to Government Budgets

A. General Approach

These stochastic population projections can be used as a basis for stochastic projections of government budgets. The Congressional Budget Office (CBO) has drawn on the Lee-Tuljapurkar set of stochastic simulations to construct long term probabilistic projections of the federal budget (CBO, 1996, 1997, 1998). Over the past few years, Lee and Tuljapurkar have developed their own set of stochastic budget projections, with a special emphasis on Social Security (Lee and Tuljapurkar, 1998a and b), but more recently including forecasts for government budgets in general (Lee, Tuljapurkar, and Edwards, 1998). Here we will give a few illustrations, and sketch the general approach.

In addition to the stochastic demography, which has already been described, we develop simple time series models for real productivity growth (purged of age composition effects), real interest rates (return on Treasury Bills held by Social Security), and, in some applications, the return on stock market equity. In all cases, we use constrained-mean models. For most of the work, productivity growth rates and interest rates are modeled as independent time series, but when the return on equity is included, a VAR is used. Future work will seek to treat inflation and unemployment in similar ways. We do not incorporate any feedbacks from the outcomes of our budget projections to these economic variables. It may seem odd to treat only four or five variables as stochastic, while treating many other important and uncertain variables, such as health care costs per enrollee, future disability rates, or defense expenditures as if their future trajectories were known with certainty. The structure of transfer programs is constantly changing, yet we take these as frozen for the next 75 years by current legislation. We do not have a good response to this objection. Perhaps these stochastic forecasts should be taken as indicating a lower bound on uncertainty about the future. Perhaps they should be viewed as forecasts conditional on some assumptions known to be unrealistic, yet capable of shedding light on the long term viability of those assumptions.

Using the Current Population Survey and published Social Security Administrative data, we have estimated average cross sectional age profiles of tax payments (for seven kinds of federal, state and local taxes) and costs of program use (for 28 kinds of federal, state and local services, including some that are quasi-public goods). In most cases, age schedules for both tax payments and costs of benefits are shifted each year with productivity growth, while debt or trust funds are projected based on government deficits or surpluses, and the interest rate (or in some experiments for Social Security, the rate of return on equities for a portion of the Trust Fund). We have generally tried to follow the approach and assumptions of the long run projections by the CBO, except in a few respects. We do forecasts under a variety of assumptions about budget balancing, for example that payroll tax rates are varied so as to keep the Social Security Trust fund equal to 100% of the following year’s expenditures; and all other tax rates are varied so as to prevent the federal debt to GDP ratio from exceeding .8.

We modify our estimated age profiles for Social Security to reflect legislated changes in the Normal Retirement Age (NRA), and experiment with further variations in the NRA. We follow the health care cost assumptions of the Medicare Trustees (Board of Trustees, 1996), which imply that the rate of increase in per enrollee costs declines gradually to the rate of productivity growth over the next few decades (see also Lee and Skinner, 1999). We depart from the Medicare Trustees’ procedures in explicitly distinguishing among the health costs of those dying within a year, those dying within 1 to 3 years, and all others, at each age (Lubitz, Beebe and Baker, 1995; Miller, 1998). This leads to lower cost projections, because declining mortality now has little effect on total health care costs: increases in the elderly population at each age are then offset by decreases in the proportion of elderly who are within three years of death at each age. As a consequence, uncertainty about the future course of mortality is largely filtered out, and the health cost projections therefore are have narrower intervals than otherwise.

B. Fiscal Implications of Demographic Change

[Figure 12 here].

Figure 12, drawn from Lee, Tuljapurkar and Edwards (1998), shows the forecasts and probability intervals for government expenditures (under currently legislated benefit structures) separately for programs that are primarily directed toward the elderly, primarily toward children, or are primarily age neutral.[15] A fourth panel shows overall expenditures, the aggregate of these other three. Focusing on central values, we see that old age expenditures will rise dramatically from 8.5% of GDP to 22.5%, reflecting both population aging and rising health costs per enrollee. Expenditures on children remain flat, as do age-neutral expenditures. Total government expenditures rise in line with the expenditures for the elderly.

The shapes of the probability fans are very different. The fan for children is very narrow for the first five years before the uncertain births reach school age, but after this the fan opens up very rapidly due to uncertain fertility, then levels off. For the elderly, the fan begins quite narrow, since only uncertain mortality is affecting the numbers of elderly. The aging of the baby boom is predictable with only moderate uncertainty, but the number of workers who will generate GDP becomes increasingly uncertain, widening the fan on expenditures as a share of GDP. Finally, once the uncertain births begin to retire, 60 or so years into the forecast, the fan widens rapidly. The age neutral forecasts are highly certain, since many are a constant share of GDP by assumption, and only the Earned Income Tax Credit is age dependent. The shape of the overall fan reflects these three components, and the covariations among them.

Analysis in Lee, Tuljapurkar, and Edwards (1998) indicates that the OASDI program accounts for less than 30% of the increase in expenditures on the elderly, with combined health care programs accounting for 57% of the increase. Fixing Social Security will not by itself solve the long run budgetary problems, although it obviously must be an important part of any solution.

C. Social Security

According to the most recent Trustees' Report (1998:23), long run actuarial balance could be restored through 2072 if the pay roll tax rate were raised by 2.19 percent. We have examined this policy option using our stochastic simulations, and found that with a 2% increase, there would still be a 75% probability of exhaustion before 2070. Policies that appear to work well in the mean often turn out to fail in the majority of cases in our simulations.

[Figure 13 here].

Here we will show results of three different kinds of stochastic forecasts, each conditional on a different assumption about policy. First consider the tax rate necessary each year to provide a reserve fund equal to 100% of the next year's expenditures on benefits, where the tax rate is left fixed at the current 12.4% until it needs to be raised to meet this condition. The resulting probability distributions for the payroll tax rate are shown in Figure 13. The median tax would remain at 12.4% until 2022, rise rapidly as the baby boom retires, and then rise more slowly to 21% in 2070, doubling from its initial value. The lower .025 probability bound remains at 12.4% until 2043, and then rises modestly to 15% in 2070. The upper 97.5% bound rises roughly linearly from 2003 to 34% in 2070. These projected payroll tax rates are conceptually similar to the SSA cost rate projections, for expenditures as a percent of payroll. Their High-Medium-Low cost rates for 2070 are 29%, 20%, and 14% (Board of Trustees, 1998:109), which may be compared to our 34%, 21% and 15%.

[Figure 14 here]

We next consider the effects of investing part of the trust fund in equities, with a stochastic return constrained to have a long run mean of 7% (real). Figure 14 plots histograms for date of Trust Fund exhaustion. Panel A shows that under current policy of investing 100% in Treasury Bills there is only a 2.5% chance of non-exhaustion before 2071, with a median date of exhaustion around 2031, consistent with SSA projections. Panel B shows that if 50% of the Fund were invested in equities by 2010, the median date of exhaustion moves back slightly to 2037, and the chance of non-exhaustion by 2071 rises to 16.7%. Panel C considers what would happen if 90% of the Fund were invested in equities in 2000. In a deterministic simulation of this policy, the Fund would still be positive after a century, in 2097, so this approach appears very promising. However, the stochastic simulation shows that the median date of exhaustion would be pushed back only to 2045, and there would still be a two thirds chance of Fund exhaustion by 2072.[16]

[Figure 15 here].

Our last example considers the effects of increasing the NRA. Panel A of Figure 15 shows a histogram for the long run actuarial balance (1998-2072) under current legislation (which raises the NRA to 67 by 2022). The median actuarial balance is -2.4%, not much different than the intermediate Trustees (1998:23) value of -2.19%. The 95% interval is -6.26% to -0.01%, which is quite similar to the Trustees' range of -5.4% to +.25%. Panel B shows the distribution if the NRA is increased to age 70 by 2033. The median has now moved up to -1.37%, and the 95% interval has moved up to -4.4% to +0.63%. Panel C shows the distribution if the NRA is increased to 71 years by 2022. The median actuarial balance is now +0.24%, and the 95% range is -1.79% to +1.61%. There is still a 34% probability of fund exhaustion before 2070 under this scenario, but there is also a 15% chance that the trust fund in 2072 would exceed 20 trillion dollars (of course, with this large a trust fund, the assumption of no economic feedback is unacceptable).

D. The Past Record of Projections of Social Security Finances by the Trustees

The OASSA has occasionally published evaluations of the performance of their projections, but these have been of limited scope. The most comprehensive we could find was Bayo (1990), which examines projections made from 1980 to 1989. Bayo (1990:1) argues convincingly that the cost rate (that is, program costs as a proportion of taxable payroll) is the most suitable item for ex post analysis, because it is less subject to policy changes than the others. Bayo (1990:12) concluded that projections of cost rates done in 1983 to 1990 were generally pessimistic relative to subsequent reality, which he attributes to the economic recovery during this period. We present a more comprehensive analysis of the performance of projections done between 1950 and 1989. Before 1972, the projections used a "level earnings, level benefits" method which made no attempt to take into account the effect on benefits of future productivity growth. The projections properly made no attempt to project the sporadic changes in legislated benefit. We have adjusted these projections ex post to reflect the actual benefit changes that subsequently occurred, providing a more comparable series (Lee and Miller, 1998).

[Table 4 here].

The first column of Table 4 gives average forecast errors by forecast horizon, where errors are measured as (Actual – Projected)/Actual.[17] The average errors in the first column reveal an excellent adjusted forecast performance, with no evident bias. Average percent errors are small at all horizons, and are sometimes positive and sometimes negative. More detailed analysis shows that this is generally true when forecasts done before and after 1972 are examined separately (not shown here), with the important exception that after 1972, longer term forecasts with a 15 to 24 year horizon averaged a negative error of 12.0 percent.

The second column shows root mean squared percent error of the forecasts, which generally increase with forecast horizon as one would expect, but dropping down a bit at the longest horizon. These are generally narrower than the high-low range of the OASSA projections, but of a similar magnitude. They are slightly broader than the ex ante forecast standard errors calculated from the LT stochastic simulations, as shown in the last column. This is what we would expect, because the LT forecasts by construction have the "right" long run mean values. The LT standard errors are half the width of the Trustees' High-Low range.

VIII. Conclusion

We begin by summarizing our main points, and then conclude with thoughts about the uses to which stochastic forecasts might be put.

A. Summary

1. Assessing Recent Official US Vital Rate Forecasts

- In retrospect, it appears that over the last fifty years, the Census and Social Security forecasters attached too much importance to the most recently observed levels of fertility and mortality.
- Recent Census Bureau projections of US fertility, based on race/ethnic disaggregation, appear to be too high. Recent Social Security fertility projections appear reasonable, although the range may be too narrow in light of international experience.

- Recent projections of life expectancy gains by both Census and Social Security appear to be substantially too low, in light of past US experience and international levels and trends in low mortality countries.

2. Uncertainty in Population Forecasts

- The standard method for dealing with uncertainty in demographic (and many other) forecasts is the use of high, medium and low scenarios. This approach is deeply flawed, because it is based on very strong and implausible assumptions about the correlation of forecast errors over time, and between fertility and mortality. The random scenario method is an improvement, but it retains some of the same flaws.
- Stochastic population forecasts based on time series models of vital rates (Lee-Tuljapurkar) appear to offer some important advantages, although long forecast horizons in demography far exceed the intended use of these models, and it is necessary to impose external constraints on the models in some cases to obtain plausible forecast behavior. One should not rely on mechanical time series forecasts in any case; they should be assessed in relation to external information.
- A parsimonious time series model for mortality appears to perform well within sample in applications in various countries, and suggests future life expectancy gains in the US at roughly twice the rate projected by Census and Social Security.

3. Results of Population Forecasts

- Middle forecasts by Census, Social Security, and Lee-Tuljapurkar (LT) agree closely on the timing and extent of increase in old age dependency ratios as the baby boom ages (although LT are somewhat higher due to lower mortality), but Census shows some amelioration after 2040, due to higher fertility. After 2040, LT forecasts continue to increase, doubling by 2070 to .45, while Social Security forecasts increase to .41. The Social Security range is three times as wide as that of Census, reflecting inherent flaws in the scenario method.
- Middle forecasts of the Total Dependency Ratio by Census and LT agree fairly closely , but are somewhat higher than Social Security (LT is .88 in 2070; Social Security is .83). The Social Security range is extremely narrow, reflecting inherent flaws in the scenario method, but the Census range is far too narrow as well.
- In Social Security forecasts, the correlation between errors in forecasting Youth Dependency Ratios and Old Age Dependency Ratios is close to –1.0. In Census forecasts, it is moderately positive. These correlations result from the bundling of assumptions in scenarios. LT forecasts show a correlation of -.6 to -.4, indicating partially offsetting variations in the proportions of children and elderly, as one would expect.

4. Stochastic fiscal projections:

- We did an analysis of the performance of Social Security projections of cost rates since 1950, for forecast horizons of up to 35 years. Performance was generally very good, with no systematic bias, small average errors, and root mean squared errors smaller than the published high-low ranges. Projections done from the mid-70s to the mid ‘80s have under-projected costs by 12%, however.
- Middle LT forecasts suggest that government expenditures on the elderly will increase by over 150% in relation to GDP by 2070, while expenditures on children and age neutral expenditures will remain flat. Taxes rise from 24% now to 38% of GDP in the median forecast to 2070 (if debt/GDP is constrained), while the 95% probability range for taxes in 2070 goes from 25% to 53% of GDP.
- Increased costs of OASDI account for nearly 30% of the increase in expenditures on the elderly, but a larger share, 57%, is due to health costs in the median forecast. Fixing Social Security will not take care of the long term budget problem.
- Investing 90% of the Social Security reserve fund in equities yielding 7% (real) would fix the system according to a deterministic simulation, but in a stochastic forecast there is still a two thirds chance of exhaustion, with a median exhaustion date of 2044, and a negative median (but strongly positive mean) Fund balance in 2050.
- Raising the payroll tax rate by 2% immediately should nearly put the system in long term actuarial balance according to Social Security projections, but still leaves a 75% chance of fund exhaustion before 2070 in LT stochastic forecasts.
- Raising the normal retirement age to 71 by 2023 raises the median long term actuarial balance above 0 in LT stochastic forecasts, but still leaves a 43% chance of fund exhaustion before 2070.

B. Some possible uses for stochastic population forecasts and demographically based stochastic fiscal forecasts

Once our research project had progressed to the point of producing stochastic population projections, we found that people were not sure how they differed from conventional scenario based forecasts, nor was it clear how and why they might be used. We decided that we should develop some applications ourselves, or otherwise the projections would most likely languish in academic journals. This decision led to the stochastic projections of government budgets which we have described and illustrated in this paper. Now the question arises, what use are these stochastic fiscal forecasts? How might they be put to use?

We suggest that policies be viewed as filters which attenuate or amplify the consequences of variance and uncertainty. Any calculation that can be done recursively along a deterministic economic-demographic trajectory can also be done along each of the 1000 sample paths in a stochastic simulation set. Proceeding in this way, it should be straight forward to evaluate the success of different policies in dealing with uncertainty. One of the most basic questions is the following: In the face of uncertainty about the future finances of the Social Security system, should we accumulate a large reserve fund to buffer the system against likely future adversity, or should we tailor our policy to future realities as they unfold? Of course we would like to achieve a high expected rate of return, by keeping taxes low and benefits high. But there are also specific concerns with variability. Given the nonlinearity of the deadweight loss function for taxes, there is an advantage to avoiding unnecessary changes in tax rates. Considerations of intergenerational equity argue for equalizing cohort rates of return, in the context of changing tax and benefit regimes and increasing survival in old age. Given an appropriate loss function, the success of alternative policies in handling uncertainty could be assessed. Alternatively, more formal methods might be used to calculate the optimal tax rate trajectory. Issues of this sort are considered in Auerbach and Hassett (in press).

The stochastic simulations could also be used as a testing ground for theoretical ideas developed using simpler but unrealistic models. Smetters (1998) finds that in a stochastic overlapping generations economy with two population age groups, prefunding a public defined benefit pension system like our Social Security would reduce ex ante unfunded liabilities much more than would replacing it with a system of private accounts, when in both cases funds are invested in equities. An older literature debated whether in an unfunded system, intergenerational inequities would be better smoothed by a defined contribution or a defined benefit system. Stochastic fiscal and demographic forecasts could readily be used to examine issues of this sort.

We hope that the availability of the stochastic simulations/forecasts will stimulate others to think about potential uses, and to put them to use in new ways.

NOTES

[1] If every woman in the population postponed any planned birth for a year during some calendar year, then in that year, the Total Fertility Rate (TFR) would be zero (except for accidental births). From this we can see that if 10% of women postpone their births in a year, the TFR would drop by 10%. We might suspect, then, that if the cross-sectional mean age at childbearing rises .1 years in a year, that the TFR might be artificially reduced by 10%. If each generation is actually planning on having two children, then such a change in timing could depress the TFR by .2 births, making it appear that fertility was far lower than its true underlying trend. Timing changes of this sort can continue for many years, so that the distortions in observed fertility can be persistent. In France, the TFR has been below replacement since the mid-1970s, although women have been having 2.1 children on average by the end of their reproductive years (Bongaarts, 1998).

[2] This kind of question has been asked in many different ways: expected, ideal, wanted, desired, desired if life could be lived over, desired given your actual economic circumstances, and so on. Originally only married women were sampled, but more recently the surveys have been broadened to include all women.

[3] Lee and Carter then used a second stage, in which the SVD estimates of a(x) and b(x) were taken as given from the first stage, but k(t) was re-estimated so as to yield the observed total number of deaths in conjunction with the observed population age distribution. This second step also made it possible to extend the time series estimate of k(t) to years in which only the total number of deaths, but not its breakdown by age, was available. This is useful for countries like the US, where there is a substantial lag before age specific mortality data are published. The estimation can be achieved in a single step by using weighted SVD, where the weights are the numbers of deaths at each age; this method has been developed and applied by Wilmoth, 1993), but in this case the range of k(t) cannot be extended. The second step is necessary to achieve a good fit to the number of deaths or to life expectancy, since fitting the log of the death rates gives extremely low death rates in youth the same weight as extremely high death rates in old age. Lee and Carter took a(x) to be the sample mean of ln(m(x,t)) for each x, arguing that this would best capture the underlying age-shape of mortality. However, in this case the model will not give the best possible fit to ln(m(x,t)) for the jump-off period of the forecast. One can alternatively take a(x) to equal the log of the death rates for the last period observed, which guarantees a perfect fit to the age structure of mortality at that point. Bell (1997) has shown that this gives better forecast performance, at least over the short run.

[4] That is, it accounts for 97.5% of the variance that remains after subtracting the age specific means a(x). The a(x) factors by themselves would account for a very

high fraction of the variance in the ln(m(x,t)) matrix, because average death rates vary by a factor of 500 or 1000 across age groups in a given year.

[5] Indirect methods were used to fit the mortality model for years 1900 to 1932, before all states belonged to the death registration area. Because of uncertainty about the quality of the age specific data that has been produced for these years, the use of the indirect method, choosing k to match the number of deaths observed in each year, given the population age distribution, seems preferable.

[6] In 1918, the epidemic reduced life expectancy by 7 years, but the same increase in k in 1996 would reduce e0 by only 2.4 years.

[7] The 1997 value is not yet available. The value used here for 1997 was estimated by a regression of e0 on the age adjusted death rate in recent years, together with the observed age adjusted death rate for 1997 which is available.

[8] Accurate forecasts would have been generated had a(x) been estimated at the most recently observed values of ln(m(x,t)), instead of at the sample means (see Bell, 1997). Lee-Carter preferred to use sample means for a(x) on the grounds that they might better reflect long run age patterns in the future.

[9] Another special feature is that the TFR cannot be negative, nor can it exceed 16 or so children per woman, for biological reasons. In practice, it appears highly unlikely that the TFR could again rise above a level of, say, 4 births per woman in the US. Lee (1993) imposed limits like these on the model through logistic transformations of the data series, and re-transformations of the forecast. This approach appeared successful, but subsequent analysis by Tuljapurkar found that the probability distribution of the fertility forecasts was rectangularized, raising its variance. In subsequent work, the constrained mean alone was used. In the forecasts reported in this paper, the occasional simulated age specific fertility values below 0 was set to 0, and runs with any TFR >4 or <0 were discarded. Only about 1.5% of runs were affected.

[10] Some analysts speak of a “Second Demographic Transition” to the current very low levels of fertility observed in parts of Europe. This idea further complicates the question of whether stationary time series models are appropriate.

[11] Note that the forecasts (expected values and probability bounds) of fertility and mortality from the time series models are not themselves used in the simulations. Instead, many random sample paths are generated, describing a large number of possible trajectories.

[12] When the TFR falls below .7, negative fertility rates occur at some older ages. These are set equal to zero.

[13] Evidently, the strong autocorrelation in the fertility series is quite attenuated after 30 years.

[14] By construction, the LT forecasts have the correct long term mean, whereas this is not true of the BC or OASSA forecasts. For this and other reasons, we should expect narrower intervals for the LT forecasts than would be estimated from ex post analysis of BC or OASSA forecast performance.

[15] The programs have been grouped by age according to the average dollar weighted age of program beneficiaries. Lee, Tuljapurkar and Edwards (1998) also carries out a similar analysis, but strictly separating out expenditures at each age rather than treating whole programs as the unit. The results are very similar.

[16] The fund balance implied by the deterministic scenario with 90% of the trust fund invested in equities rises to 40% of GDP by 2050. The distributions of fund balances are highly skewed in stochastic simulations, with means far exceeding medians. For example, in 2050 the mean fund to GDP ratio is 85%, while the median is –8%.

[17] Bayo (1990) estimates errors as (Actual – Projected)/(Taxable Earnings) and concludes the projections have been successful because the errors were generally less than .8%. However, since Costs were around 11% over this period, an error of .8% is better viewed as a 7% error (.07 = .8/11), which is less satisfactory.

References

Alho, Juha M. 1990. Stochastic Methods in Population Forecasting. *International Journal of Forecasting* 6:521-530.

Alho, Juha M., and Bruce D. Spencer. 1985. Uncertain Population Forecasting. *Journal of the American Statistical Association* 80:306-314.

Alho, Juha M., and Bruce D. Spencer. 1990. Error Models for Official Mortality Forecasts. *Journal of the American Statistical Association* 85:609-616.

Auerbach, Alan and Kevin Hassett. In this volume. Uncertainty and the Design of Long-Run Fiscal Policy.

Bayo, Francisco R. 1990. *A Simple Evaluation of Projections in OASDI Trustees Reports*. Actuarial Note 132 of the Social Security Administration, US Department of Health and Human Services.

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

Board of Trustees, The Federal Hospital Insurance Trust Fund. 1996. *The 1996 Annual Report of the Board of Trustees of the Federal Hospital Insurance Trust Fund*. 104th Congress, 2nd Session, Report No. 104-227. Washington, DC: U.S. Government Printing Office.

Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. 1998. *The 1998 Annual Report*. Washington DC: U.S. Government Printing Office.

Bongaarts, John. 1998. Global Population Growth: Demographic Consequences of Declining Fertility. *Science* 282: 419-420.

Bongaarts, John and Griffith Feeney. 1998. On the Quantum and Tempo of Fertility. *Population and Development Review* 24: 271-292.

Cohen, Joel. 1986. Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-Based and Empirical Approaches. *Demography* 23:105-126.

Congressional Budget Office of the United States. 1996. *The Economic and Budget Outlook: Fiscal Years 1997-2006*. Washington, DC: U.S. Government Printing Office.

Congressional Budget Office. 1997. *Long-Term Budgetary Pressures and Policy Options*. . Washington, DC: U.S. Government Printing Office.

Congressional Budget Office. 1998. *Long-Term Budgetary Pressures and Policy Options*. . Washington, DC: U.S. Government Printing Office.

Cutler, David, James Poterba, Louise Sheiner, and Lawrence Summers. 1990. An Aging Society: Opportunity or Challenge? *Brookings Papers on Economic Activity*. 1:1-56,71-73.

Day, Jennifer Cheeseman. 1993. *Population Projections of the United States, by Age, Sex, Race, and Hispanic Origin: 1992 to 2050*. U.S. Bureau of the Census, Current Population Reports, P25-1092. Washington, DC: U.S. Government Printing Office.

Day, Jennifer Cheeseman. 1996. *Population Projections of the United States, by Age, Sex, Race, and Hispanic Origin: 1995 to 2050*. U.S. Bureau of the Census, Current Population Reports, P25-1130. Washington, DC: U.S. Government Printing Office.

Easterlin, Richard A. 1968. *Population, Labor Force, and Long Swings in Economic Growth: The American Experience*. New York: National Bureau of Economic Research.

Easterlin, Richard. 1978. What Will 1984 Be Like? Socioeconomic Implications of Recent Twists in Age Structure. *Demography* 15:397-432.

Foster, R. 1997. *Evaluation of mortality assumptions in the Trustees' report.* HCFA Memorandum. Office of the Chief Actuary, Health Care Financing Administration, Baltimore, MD.

Goldstein, Joshua. 1997. *Forecasting the future of U.S. divorce rates*. Paper presented at the 1997 Annual Meetings of the Population Association of America, Washington DC.

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.

Horiuchi, Shiro and John Wilmoth. 1995. *Aging of Mortality Decline*. Paper presented at the 1995 Annual Meetings of the Population Association of America, in San Francisco.

Kannisto, Vaino, Jens Lauritsen, A. Roger Thatcher and James Vaupel. 1994. Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries *Population and Development Review* 20:793-810.

Keyfitz, Nathan. 1981. The Limits of Population Forecasting. *Population and Development Review* 7:579-593.

Lee, Ronald. 1974. Forecasting Births in Post-Transitional Populations: Stochastic Renewal with Serially Correlated Fertility. *Journal of the American Statistical Association* 69:607-617.

Lee, Ronald. 1976. Demographic Forecasting and the Easterlin Hypothesis. *Population and Development Review* 2:459-468.

Lee, Ronald D. 1993. Modeling and Forecasting the Time Series of US Fertility: Age Patterns, Range, and Ultimate Level. *International Journal of Forecasting* 9:187-202.

Lee, Ronald. Forthcoming. Probabilistic Approaches to Population Forecasting. in Wolfgang Lutz, James Vaupel and Dennis Ahlburg, eds, Supplement to vol. 25 of *Population and Development Review*, 1999 *Rethinking Population Projections*.

Lee, Ronald D. and Lawrence Carter. 1992. Modeling and Forecasting the Time Series of U.S. Mortality. *Journal of the American Statistical Association* 87:659-671.

Lee, Ronald, and Timothy Miller. 1998. *A Research Note on the Performance of Actuarial Projections of the Cost Rate: An Ex-post Analysis*. Working Paper of the Project on the Economic Demography of Interage Transfers, Department of Demography, University of California at Berkeley.

Lee, Ronald, and Jonathan Skinner. 1996. Assessing Forecasts of Mortality, Health Status, and Health Costs During Baby-Boomers' Retirement. in C. Citro. (ed.). *Modeling the Impact of Public and Private Policies on Retirement Behavior and Income*. Washington DC: National Academy of Sciences.

Lee, Ronald and Jonathan Skinner. 1999. Will Aging Baby Boomers Bust the Federal Budget? Longevity, Health Status, and Medical Costs in the Next Century. *Journal of Economic Perspectives*

Lee, Ronald and Shripad Tuljapurkar. 1994. Stochastic Population Projections for the United States: Beyond High, Medium and Low. *Journal of the American Statistical Association* 89:1175-1189.

Lee, Ronald and Shripad Tuljapurkar. 1998a. Uncertain Demographic Futures and Social Security Finances. *American Economic Review* 88:237-241.

Lee, Ronald and Shripad Tuljapurkar. 1998b. Stochastic Forecasts for Social Security. in David Wise, ed. *Frontiers in the Economics of Aging*. Chicago: University of Chicago Press.

Lee, Ronald, Shripad Tuljapurkar, and Ryan Edwards. 1998. *Uncertain Demographic Futures and Government Budgets in the US*. Unpublished manuscript of the Department of Demography, University of California at Berkeley; presented at the 1998 Annual Meetings of the Population Association of America, Chicago.

Lubitz, James, James Beebe, and Colin Baker.. 1995. Longevity and Medicare Expenditures. *The New England Journal of Medicine* 332:999-1003.

Lutz, Wolfgang, Warren Sanderson, and Sergei Scherbov. 1996. Probabilistic Population Projections Based on Expert Opinion Chapter 16 in Wolfgang Lutz, ed.. 1996. *The Future Population of the World: What Can We Assume Today*? , revised 1996 edition. London: Earthscan Publications Ltd.

Lutz, Wolfgang and Sergei Scherbov. 1998. An Expert-Based Framework for Probabilistic National Population Projections: The Example of Austria *European Journal of Population* 14:1-17.

Lutz, Wolfgang, Warren Sanderson and Sergei Scherbov. 1997. Doubling of World Population Unlikely *Nature* 387:803-805.

Manton, Kenneth G., Larry S. Corder, and Eric Stallard. 1997. Chronic Disability Trends in Elderly United States Populations: 1982-94. *Proceedings of the National Academy of Science* 94: 2593-2598.

Manton, Kenneth, Eric Stallard and H. Dennis Tolley. 1991. Limits to Human Life Expectancy: Evidence, Prospects, and Implications. *Population and Development Review*17:603-638.

Mason, Andrew. 1996. Population and Housing *Population Research and Policy Review* 15:419-435.

McNown, Robert and Andrei Rogers. 1989. Forecasting Mortality: A Parameterized Time Series Approach. *Demography* 26:645-660.

McNown, Robert and Andrei Rogers. 1992. Forecasting Cause-Specific Mortality Using Time Series Methods. *International Journal of Forecasting* 8:413-432.

Miller, Timothy. 1998. *Longevity and Medicare*. Working Paper of the Project on the Economic Demography of Interage Transfers, Department of Demography, University of California at Berkeley.

Office of the Actuary, Social Security Administration (OASSA). 1992. *Life Tables for the United States Social Security Area, 1990-2080.* Actuarial Study No. 107, SSA Pub. No. 11-11536. Baltimore: Social Security Administration.

Office of the Actuary, Social Security Administration (OASSA). 1997. *Social Security Area Population Projections: 1997*. SSA Pub. No. 12-11553. Baltimore: Social Security Administration.

Pflaumer, Peter. 1988. Confidence Intervals for Population Projections Based on Monte Carlo Methods. *International Journal of Forecasting* 4:135-142.

Sanderson, Warren. 1995. Predictability, Complexity, and Catastrophe in a Collapsible Model of Population, Development, and Environmental Interactions in Andre Rogers, ed., special issue of *Mathematical Population Studies* 5: 259-280.

Smetters, Kent. 1998. *Privatizing Versus Prefunding Social Security in a Stochastic Economy*. Unpublished manuscript of the University of Pennsylvania, Department of Economics.

Smith, James and Barry Edmonston. 1997. *The Face of the U.S. Population in 2050*. Working paper.

Stoto, Michael. 1983. The Accuracy of Population Projections. *Journal of the American Statistical Association* 78:13-20.

Tuljapurkar, Shripad and Carl Boe. 1997. *Validation, Probability-Weighted Priors, and Information in Stochastic Forecasts*. Unpublished manuscript of Mountain View Research, Mountain View, CA.

Tuljapurkar, Shripad and Carl Boe. 1998. Mortality Change and Forecasting: How Much and How Little Do We Know? *North American Actuarial Journal* 2:3-47.

Wachter, Kenneth W. 1991. Elusive Cycles: Are There Dynamically Possible Lee-Easterlin Models for U.S. Births? *Population Studies* 45: 109-135.

Wachter, Kenneth W.. 1997. Kinship Resources for the Elderly. *Philosophical Transactions of the Royal Society B.* 352:1811-1817.

Wachter, Kenneth W. and Caleb E. Finch, eds. 1997. *Between Zeus and the Salmon: The Biodemography of Longevity*. Washington, DC: National Academy of Sciences 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.. 1995. Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* 5:293-319

Wilmoth, John R.. 1996. Mortality projections for Japan: A comparison of four methods. In: Graziella Caselli and Alan Lopez (eds.), *Health and Mortality among Elderly Populations*, Oxford University Press.

Wilmoth, John R. 1998. The future of human longevity: A demographer's perspective. *Science* 280: 395-397.

Table 1. Average Annual Rate of Decline in Mortality for Base Period versus Forecast, by Age and Sex, % per Year

Age Group	*1900-1991 (base period)*	*1995-2080 (forecast period)*	*Base rate - Forecast rate*	*Ratio of forecast to trend extrapolation in 2080*
0-14	3.27	1.52	1.75	4.4
15-64	1.39	0.68	.71	1.8
65-84	.86	0.50	.36	1.4
85+	0.54	0.49	.05	1.0
Total	1.18	0.57	-0.61	1.67

Note: Calculated from data in Office of the Actuary (1996:8-14). Rates are unweighted averages across age and sex. The last column indicates the ratio of the trend extrapolated rates to the SSA projected rates at each age and overall. It is calculated by exponentiating 85 years times .01 times the previous column.

Table 2. High-Low Ranges for Forecasts of Selected Items to 2050, as Percent of Middle Forecast (Calculated as 100(H-L)/(2*M))

	BC (‘92 forc, to 2050)	**OASSA (’92 forc, to 2050)**	**Lee-Tulja (’94 forc, to 2050)**
Children	±44	±31	±49
Working Age	±26	±13	n.a.
Elderly	±27	±9	±10
Old Age Dep Ratio 65+/ 20-64	±3	±21	±35
Tot Dep Rat (<20+65+)/ 20-64	±10	±0	±24

Calculated as the (high – low)/(2*middle). For Census, High minus Low; for Actuary, High Cost minus Low Cost; for Lee-Tulja, upper 95% bound minus lower 95% bound. The date of publication of the forecast is indicated; all are for the year 2050, which is the latest published by the Census Bureau. For Census, Children are <18; for others, <20. Elderly are always 65+. Lee and Tuljapurkar (1994) did not publish a probability bound for the working age population, so none is shown.

Table 3. Standard Errors for Forecasts of the Population Growth Rate by Horizon (percent per year)

Forecast Horizon (years)	**Census by Day (1996)**	**Census by Stoto (1983)**	**Lee-Tulja (stoch sim)**
5	0.17	0.38	0.13
10	0.31	0.48	0.17
15	0.40	0.61	0.18
20	0.45	0.65	0.19

Note: Calculations by Day (1996:30) and Stoto (1983) are ex post analyses of the difference between the projected and actual average population growth rate for each horizon, based on forecasts from 1945 to 1970 (Stoto) or 1950 to 1992 (Day). The Lee-Tuljapurkar figures are based on the standard error of projected growth rates on individual sample paths, about the mean forecast.

Table 4. Errors of Cost Rate Forecasts from Ex Post Analysis of OASSA Projections, 1950-1989, and Lee-Tuljapurkar Stochastic Forecasts (ex ante), by Horizon (Adjusted for unanticipated changes in legislated benefits)

Years since Forecast	**Average % Error (SSA:1950-89)**	**Root Mean Sqd % Error (SSA:1950-89)**	**SSA Forecast Range (%) (H-L)/(2M)**	**Stand Error (%) of Lee-Tulja Forecast**
1-4	-2.6 (7)	3.0	2.9	2.4
5-14	+0.2 (25)	7.0	11.2	5.5
15-24	-1.2 (26)	11.1	13.3	6.8
25-34	+3.6 (16)	9.1	16.9	7.9

Note: The first column is calculated from all available Reports of the Trustees, from 1950 to the present. Numbers in parentheses represent the number of observations on which the average is based; this number is the same for the next column as well. The OASSA forecasts have been adjusted to reflect subsequent unanticipated changes in legislated benefit levels, modifying the pre-1972 “level earnings, level benefit” forecasts. The SSA Forecast Range is calculated from Trustees of SSA (1998:169-170) as the 100*(High – Low)/(2*Intermediate). The Lee-Tuljapurkar standard errors are calculated from the stochastic simulations. Throughout, errors are calculated as (Actual – Forecast)/Actual.

Figure 1. TFR: Actual and Middle Series Projections 1940-2005

(Source: U.S. Census Bureau, Current Population Reports, Series P-25)

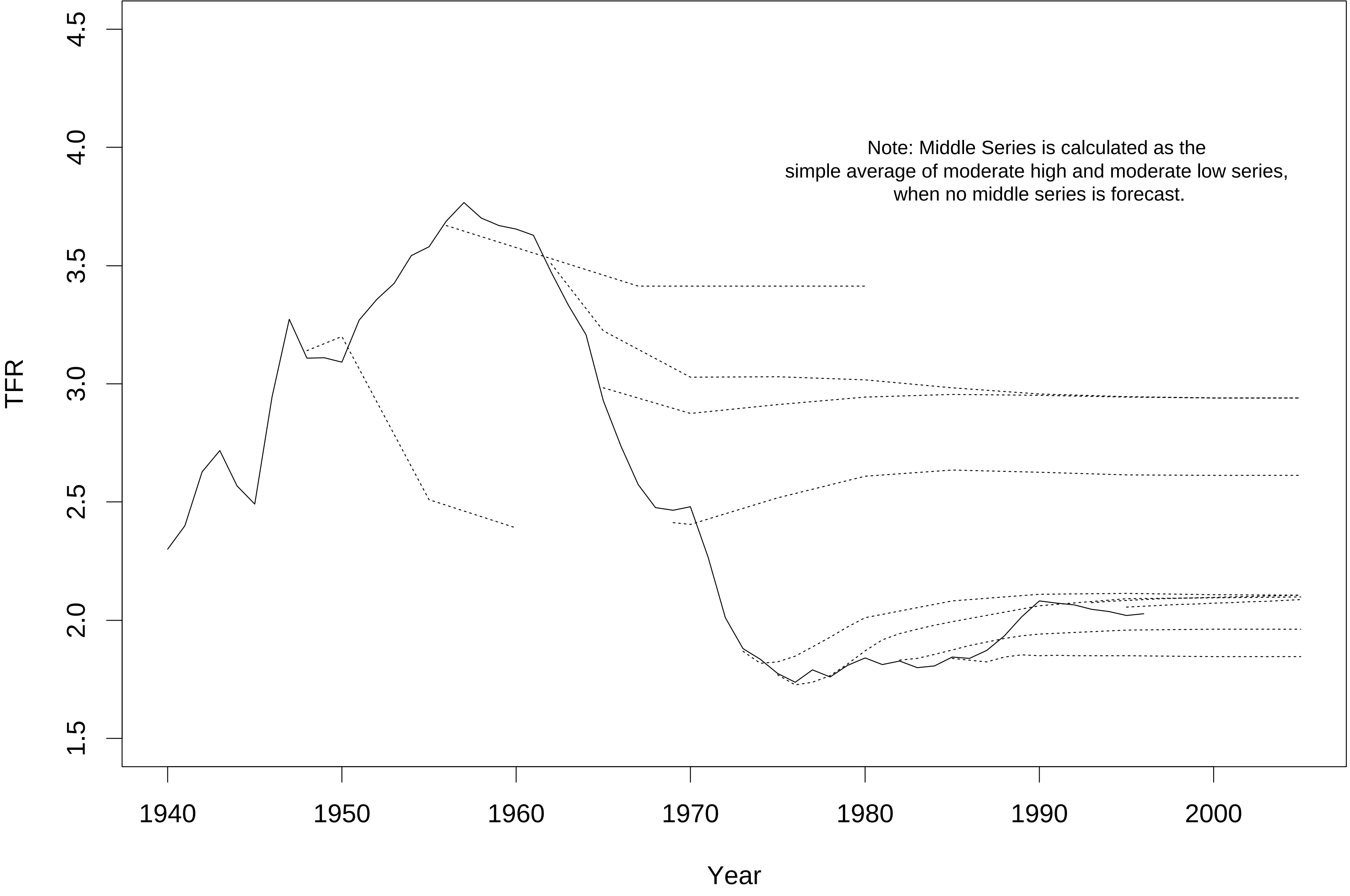


Figure 2. TFR: Actual and High/Low Series Projections 1940-2005

(Source: U.S. Census Bureau, Current Population Reports, Series P-25)

Figure 3
Actual and Middle Series Projections of
Life Expectancies for the United States
[Source: Social Security Administration, Actuarial Studies]

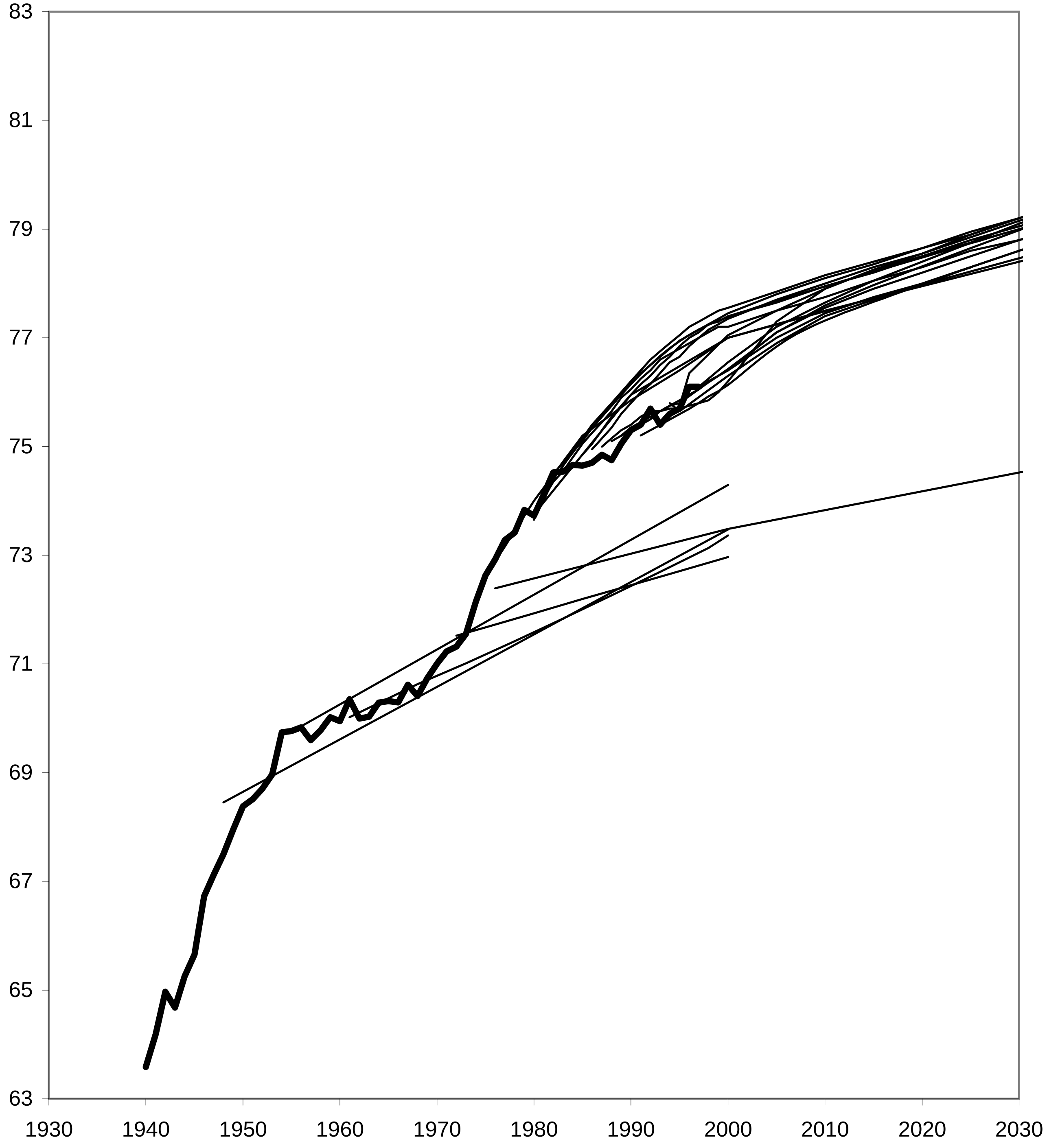
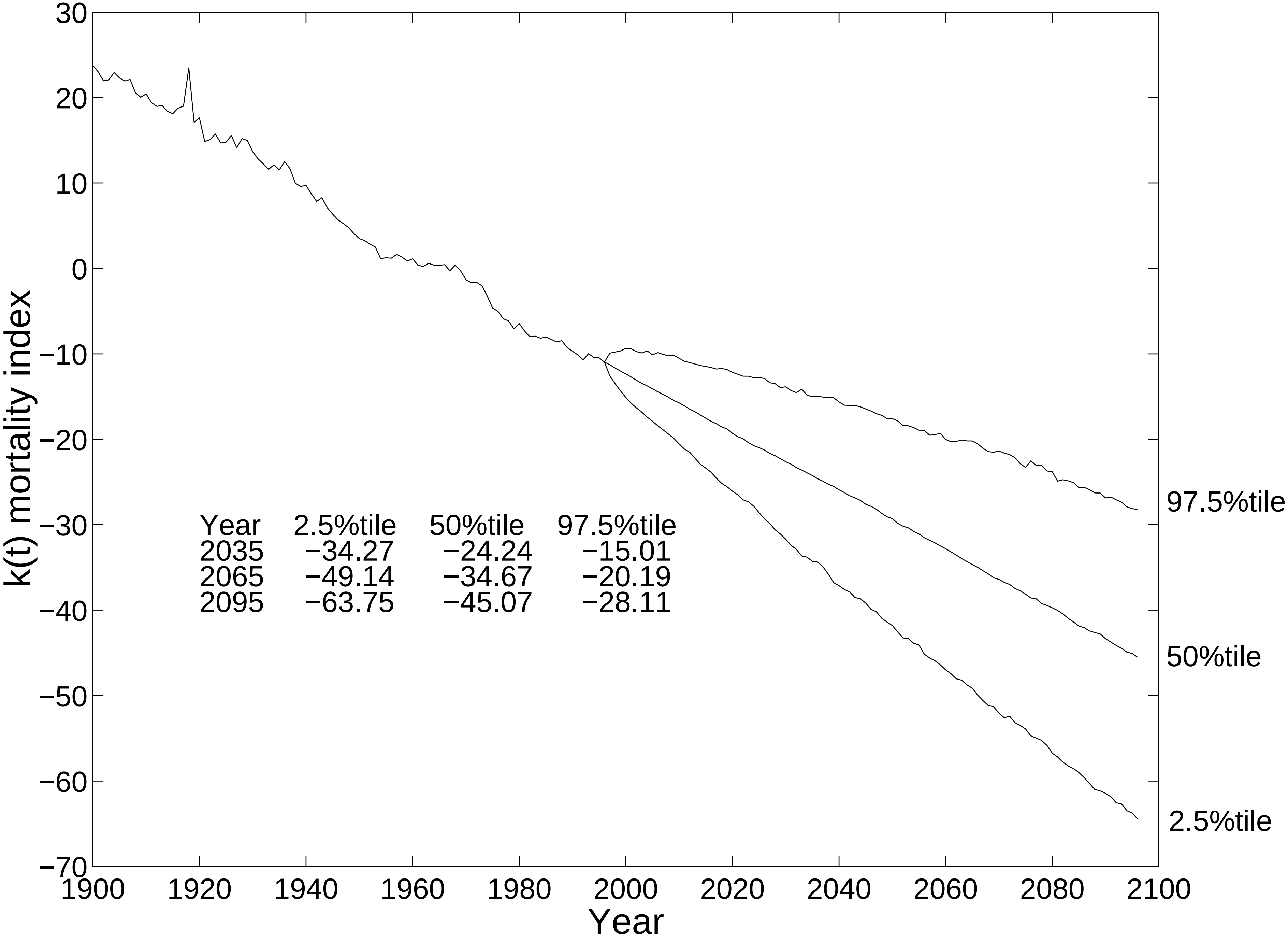
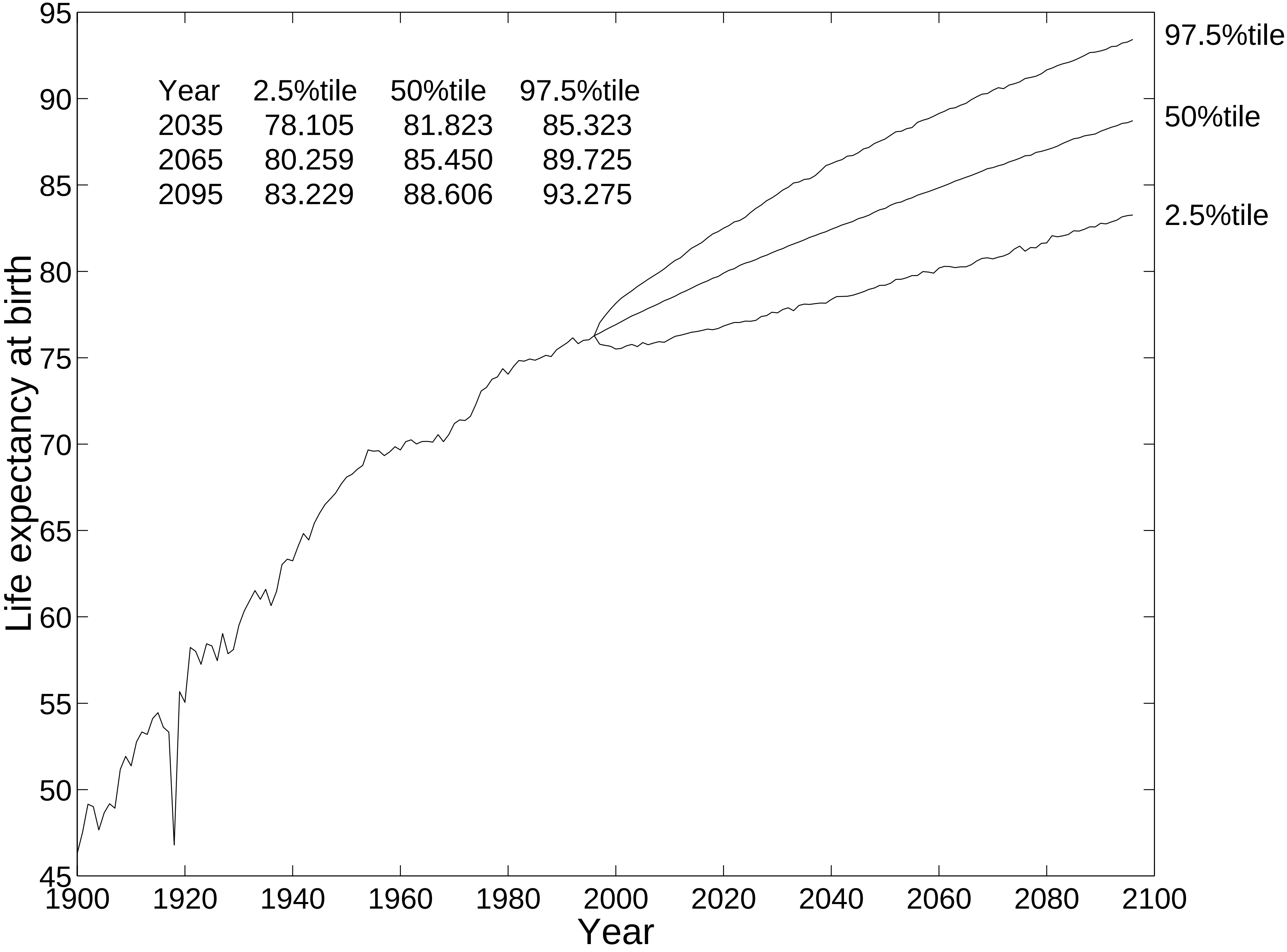


Fig. 4 − Lee−Carter mortality index k(t), fitted (1900−96) and forecasted (1997−2096)



Year	2.5%tile	50%tile	97.5%tile
2035	−34.27	−24.24	−15.01
2065	−49.14	−34.67	−20.19
2095	−63.75	−45.07	−28.11

Fig. 5 − Life expectancy at birth, fitted (1900−96) and forecasted (1997−2096)



Year	2.5%tile	50%tile	97.5%tile
2035	78.105	81.823	85.323
2065	80.259	85.450	89.725
2095	83.229	88.606	93.275

Figure 6. Projections of Life Expectancy at Birth, 1990-2000

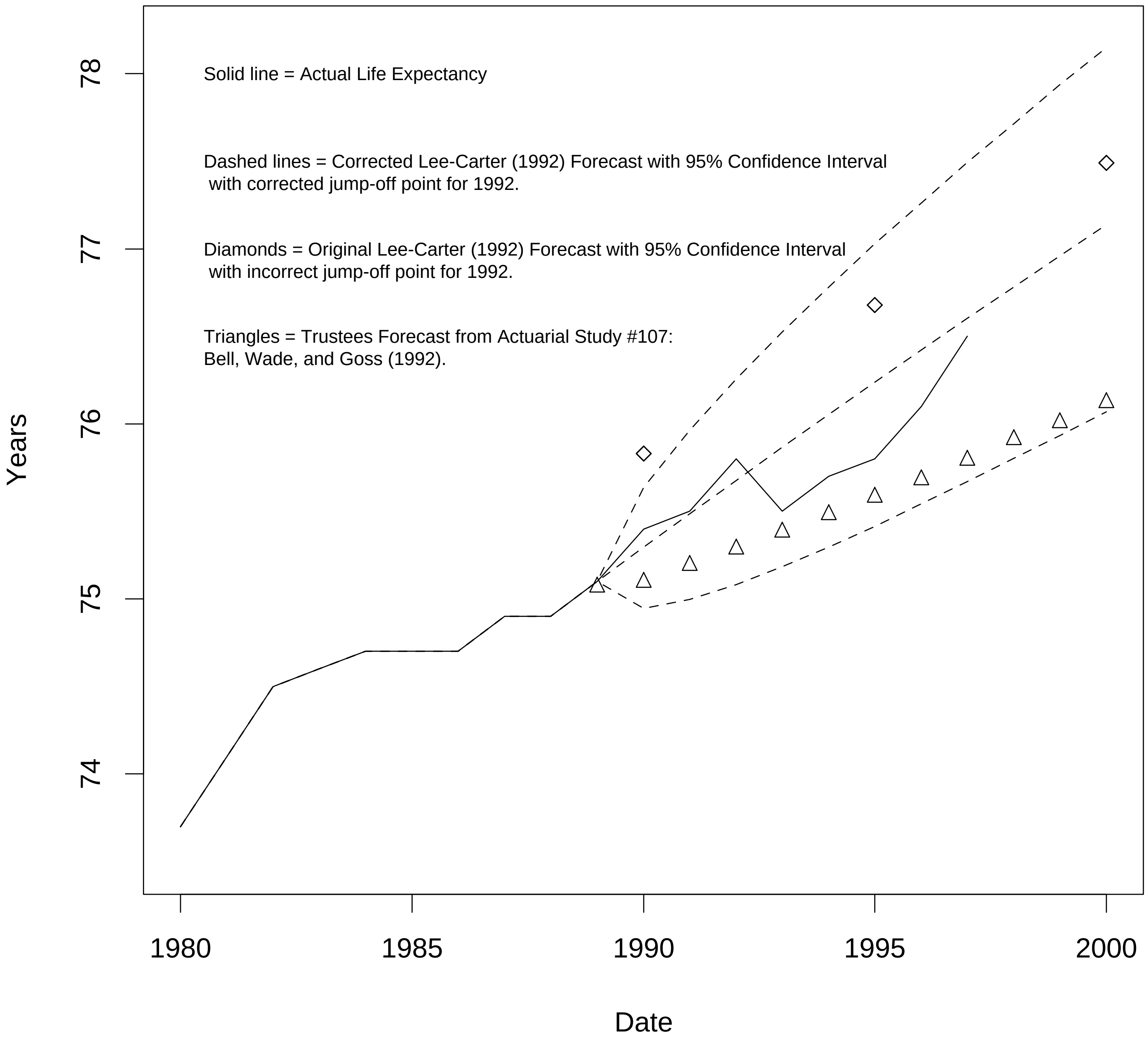
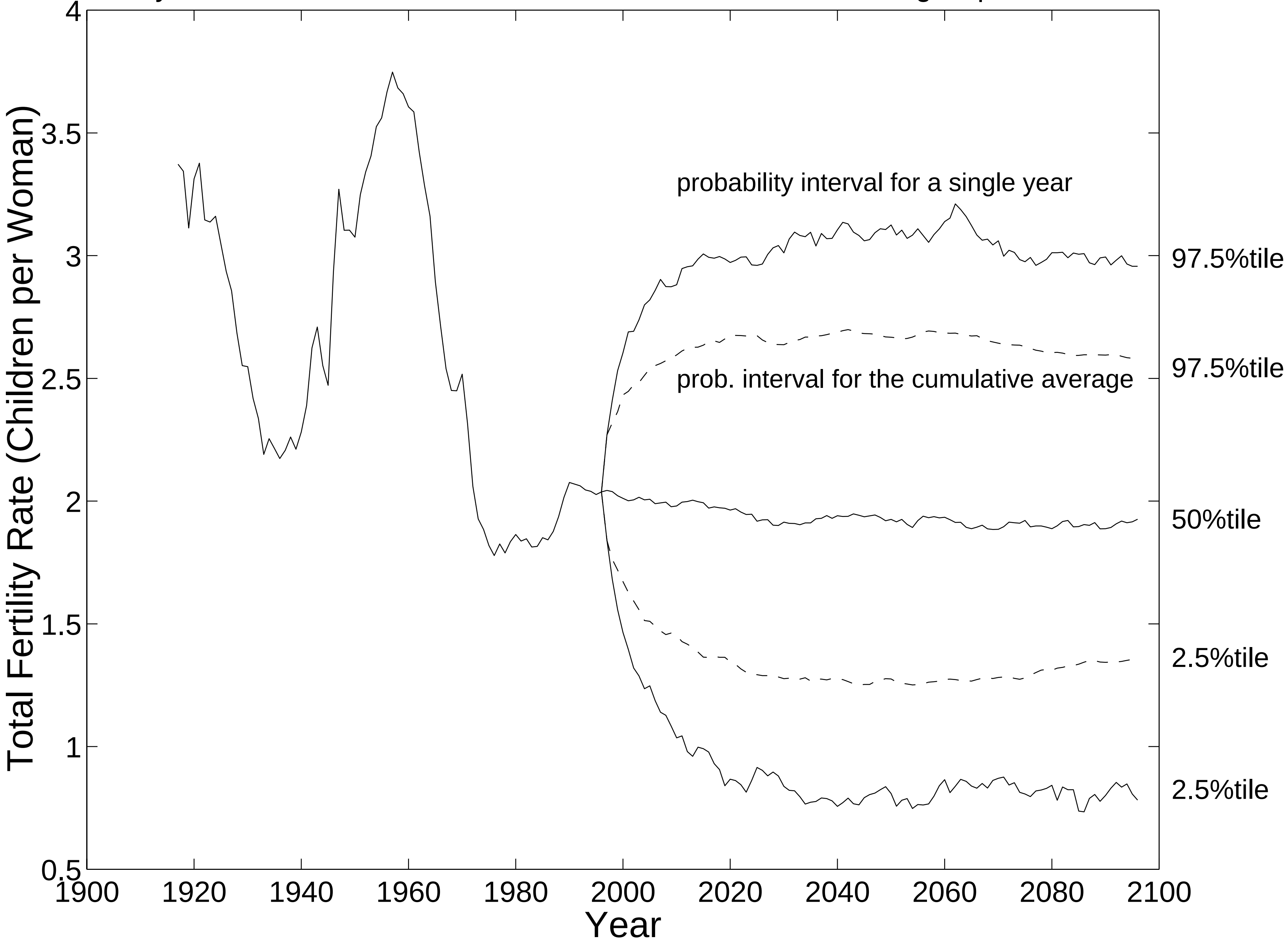


Figure 7. Total Fertility Rate, historical (1917−1996) and forecasted (1997−2096), with 95% Probability Intervals for Annual Values and for the Cumulative Average Up To Each Horizon



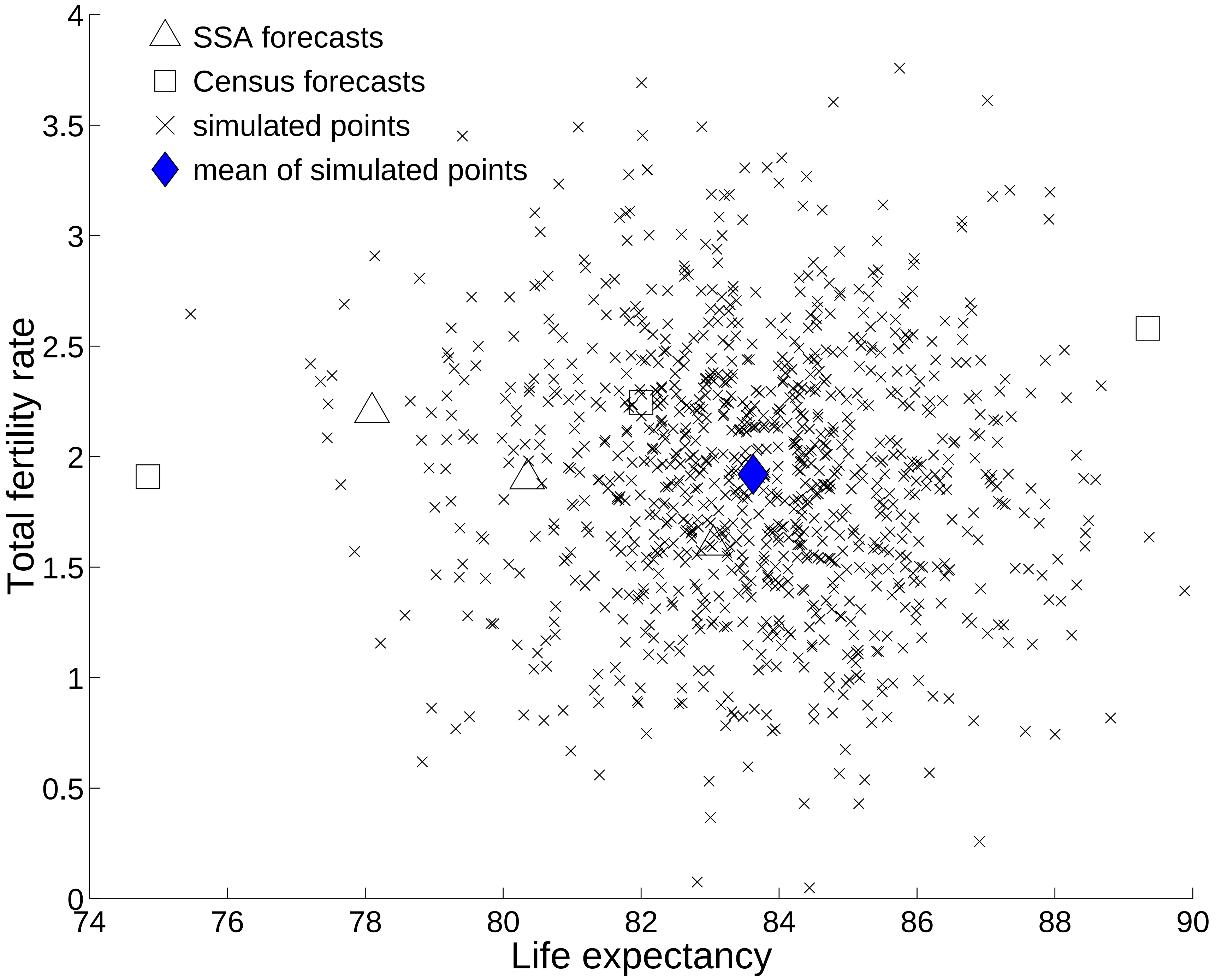


Fig. 8 − TFR by life expectancy in 2050, 1,000 simulated points

Figure 9. TFR in 2050 by TFR in 2020, 1,000 simulated points

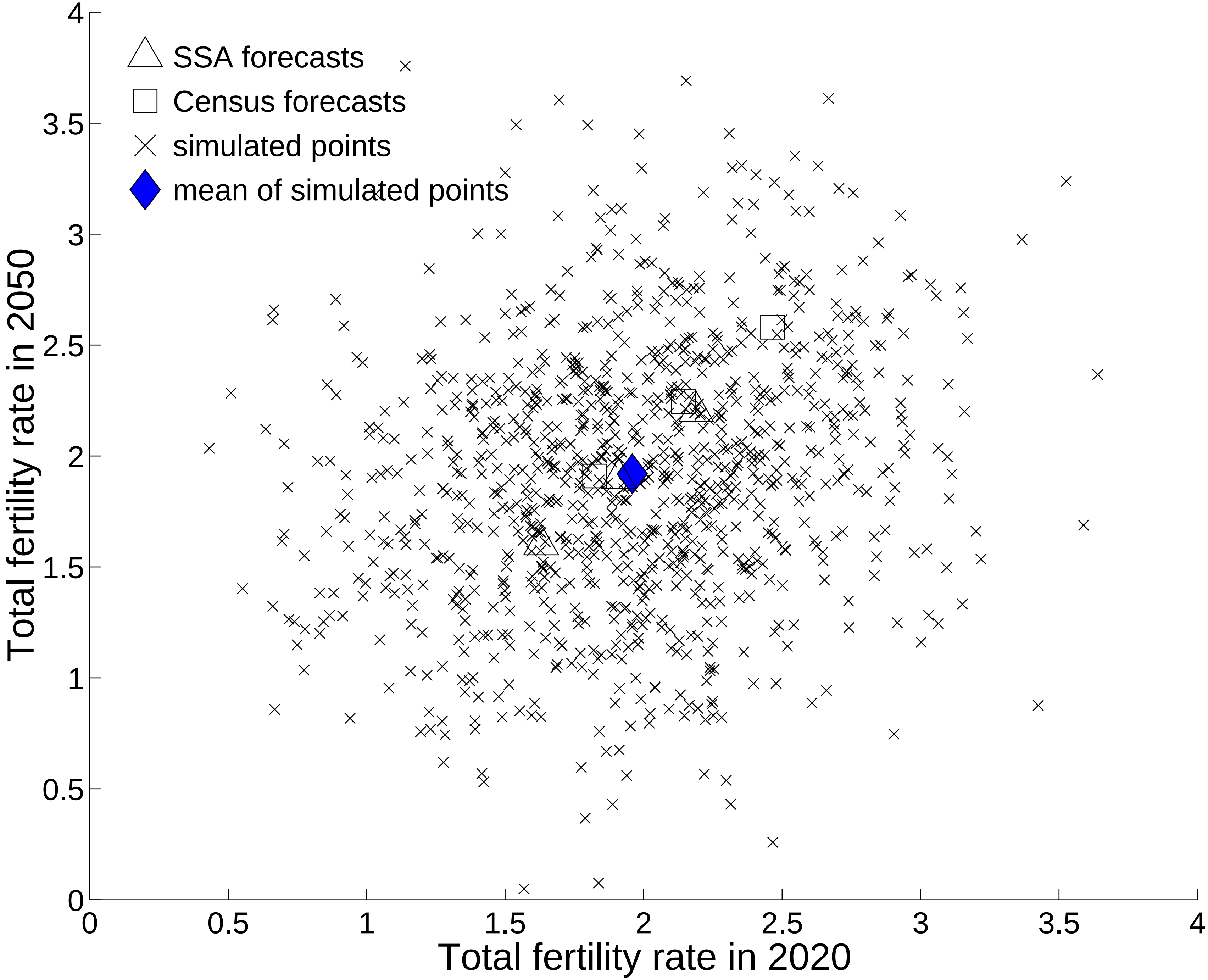


Fig. 10 − Old age dependency ratio (65+ pop)/(20−64 pop)

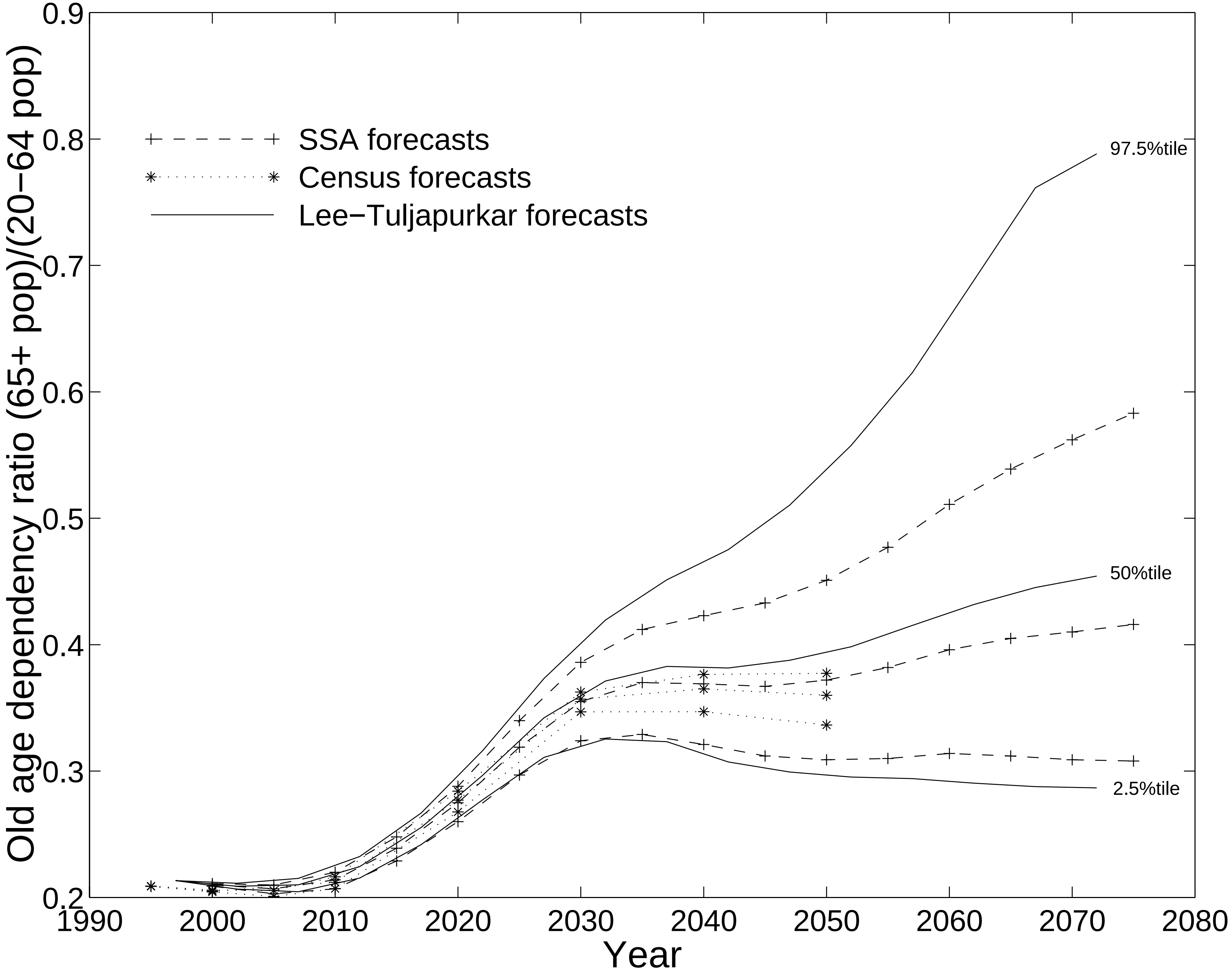
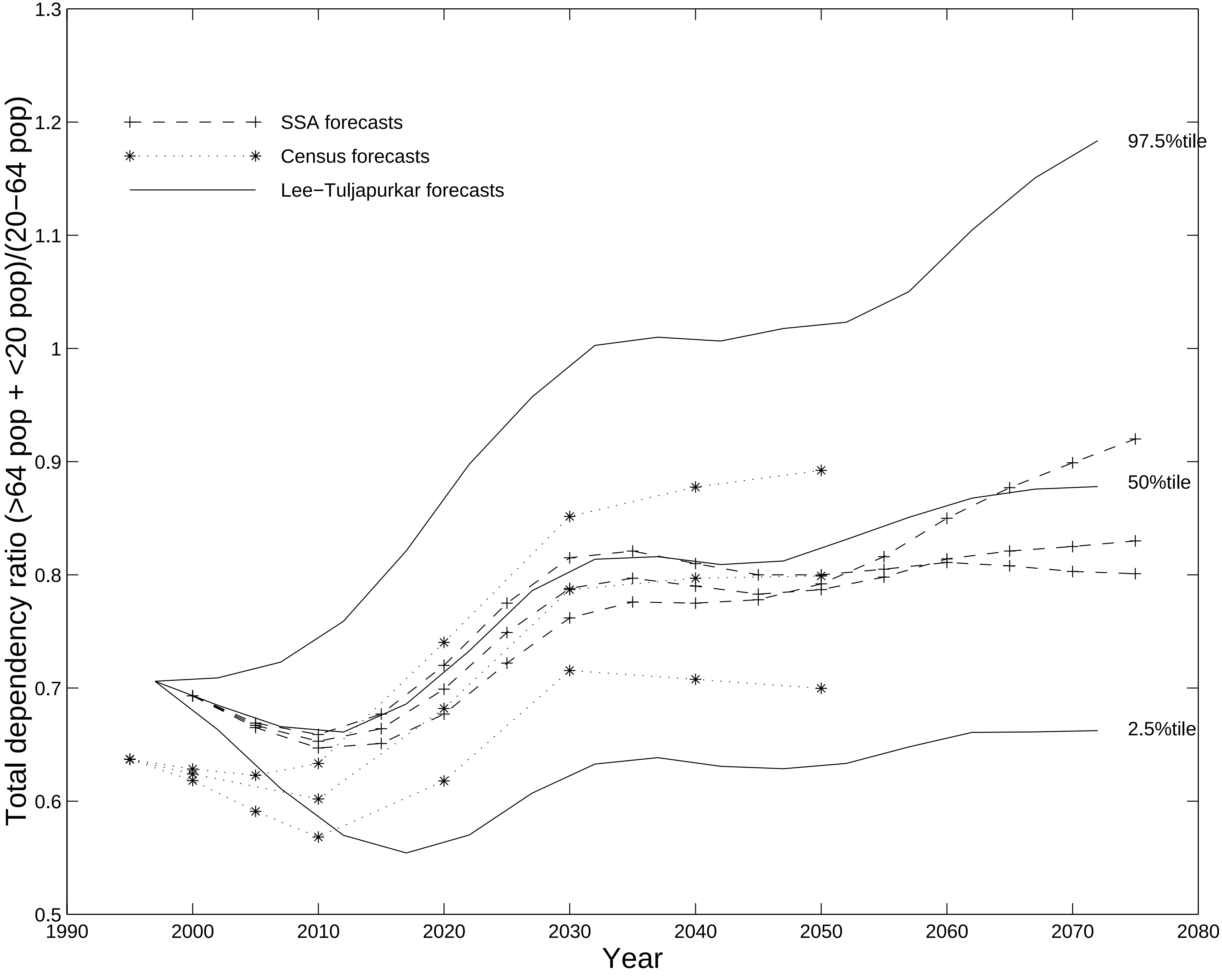
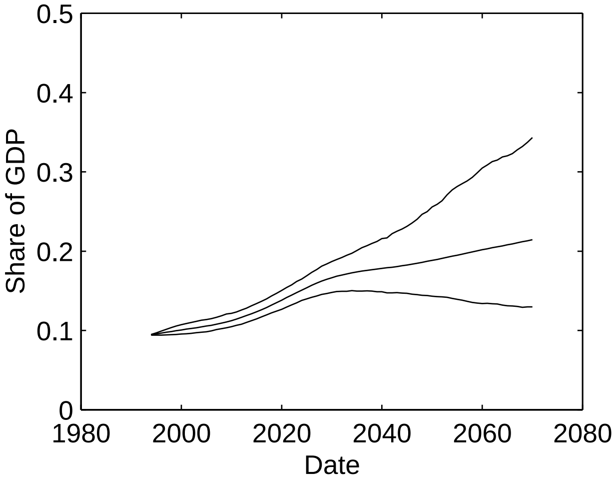


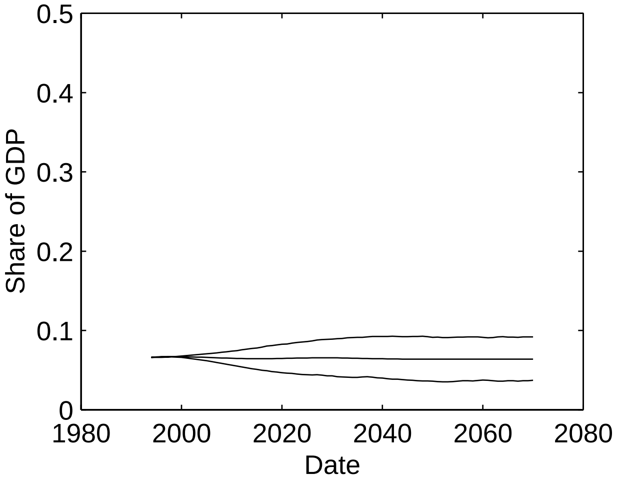
Fig. 11 − Total dependency ratio (>64 pop + <20 pop)/(20−64 pop)



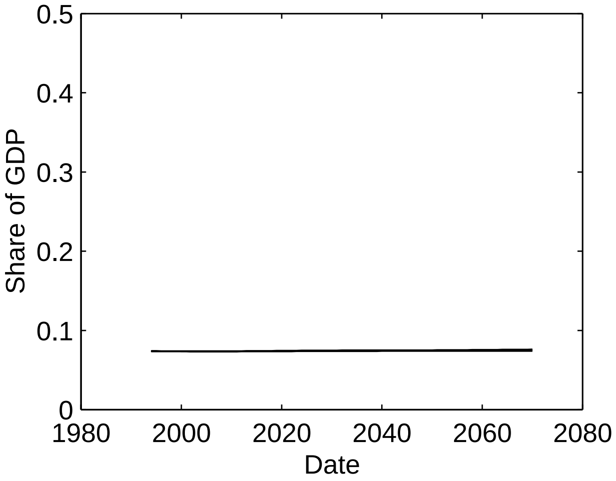
Note: Census uses (>64 pop + <18 pop)/(18−64 pop).



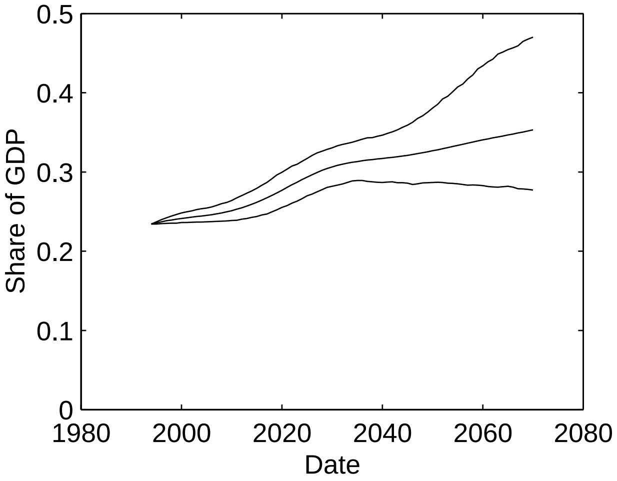
12A. Spending on Programs for Old



12B. Spending on Programs for Young



12C. Spending on Age−Neutral Programs



12D. Spending on All Government Programs

Figure 13. Tax Rate for Year-Ahead Balance
(Distribution by Selected Probability Percentiles)

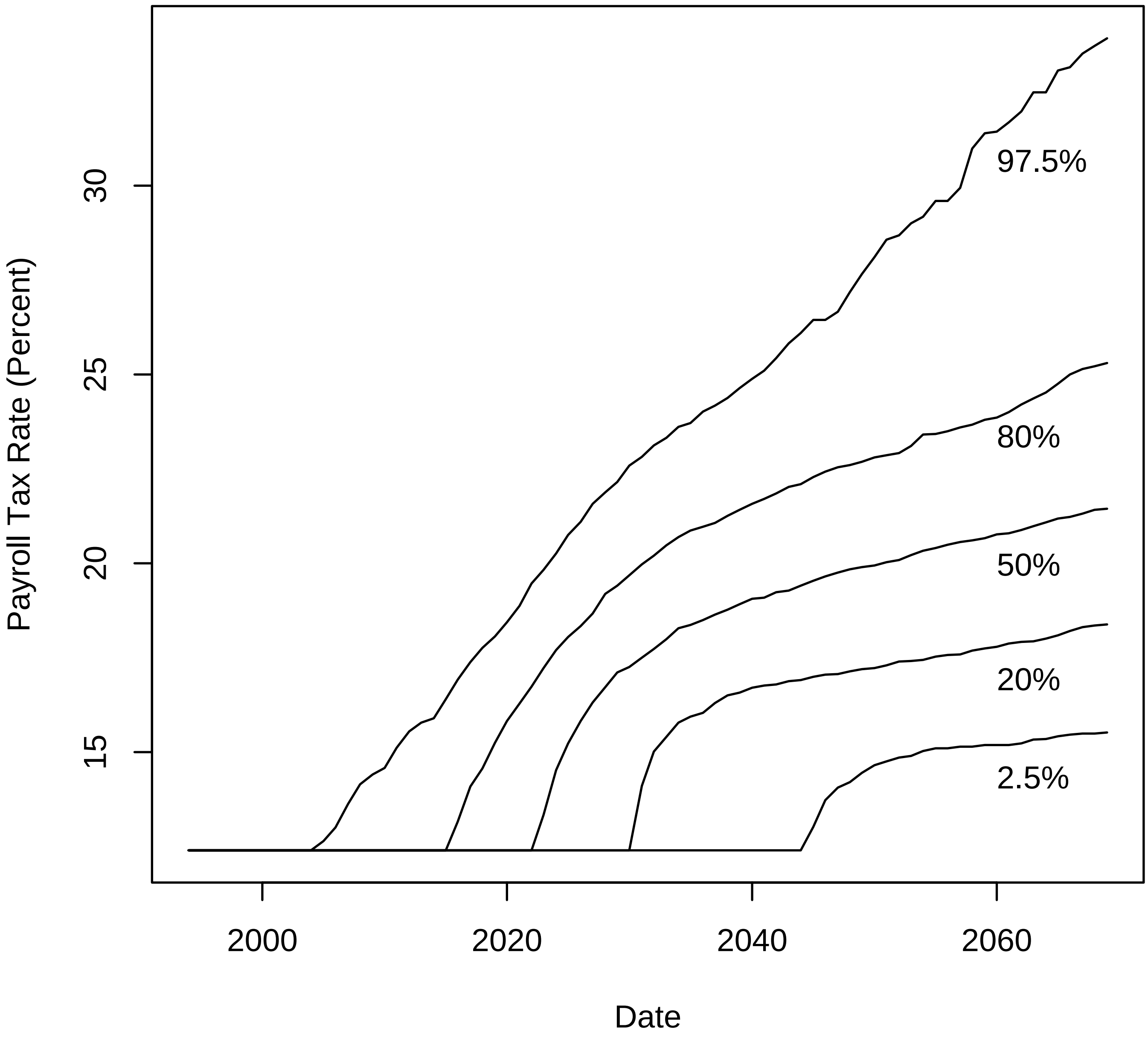
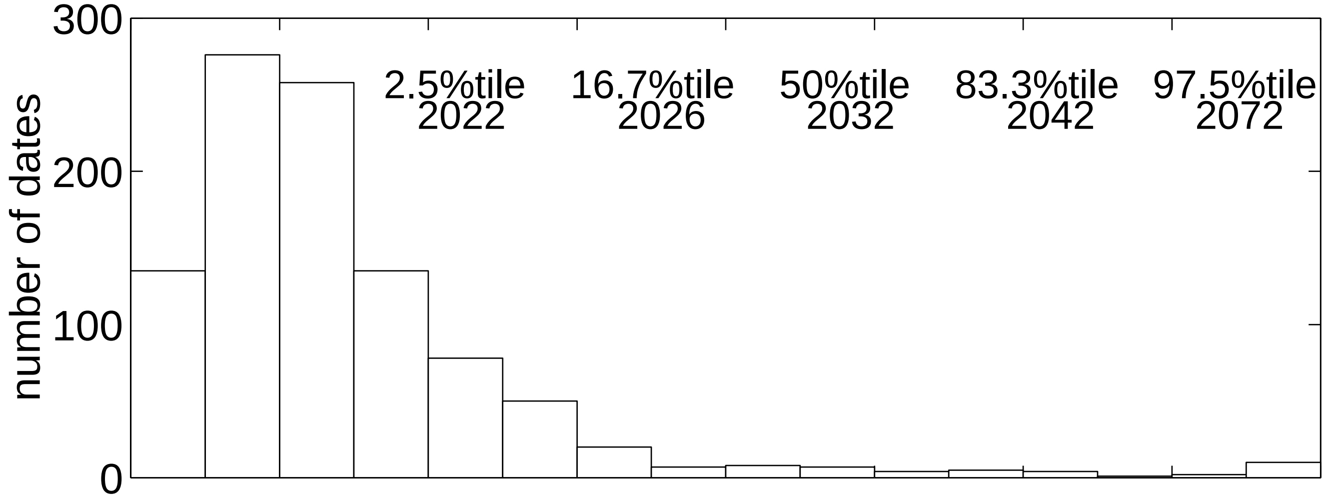
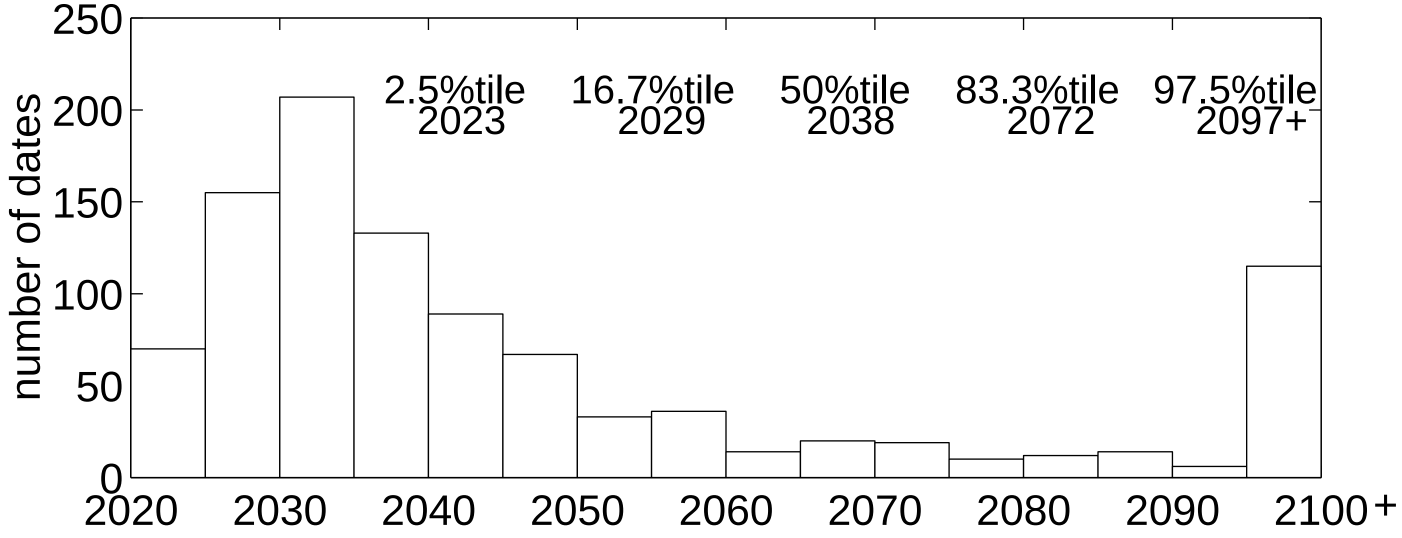


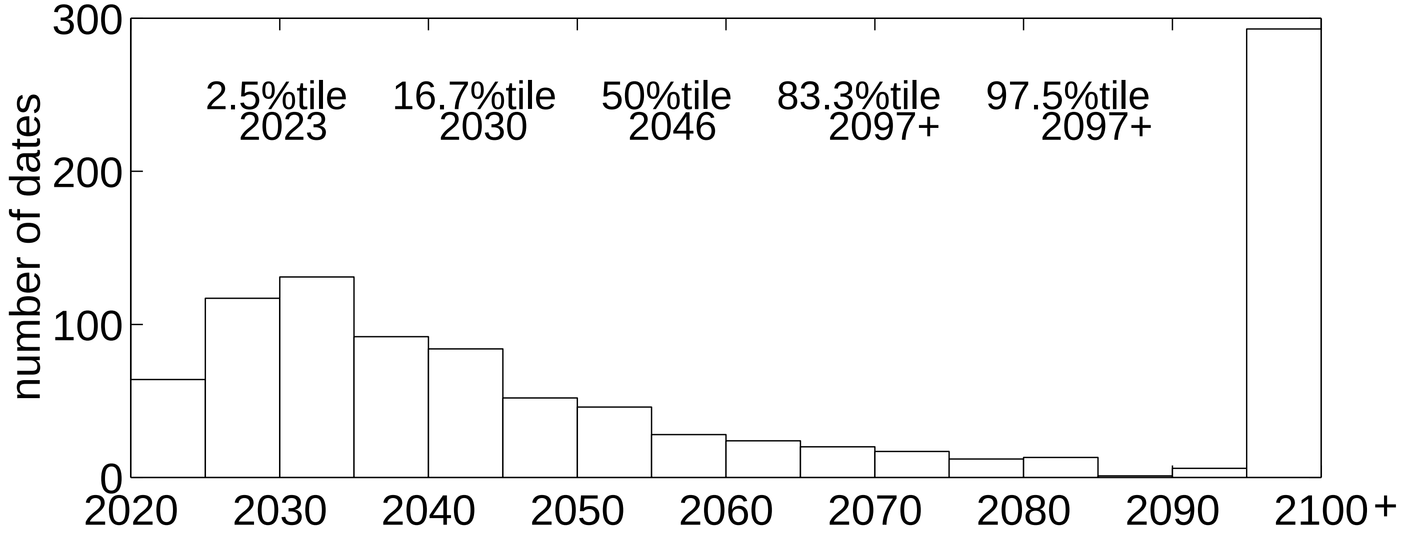
Fig. 14 − Histograms of 1,000 dates of exhaustion for three scenarios



Panel A − No investment in S&P 500

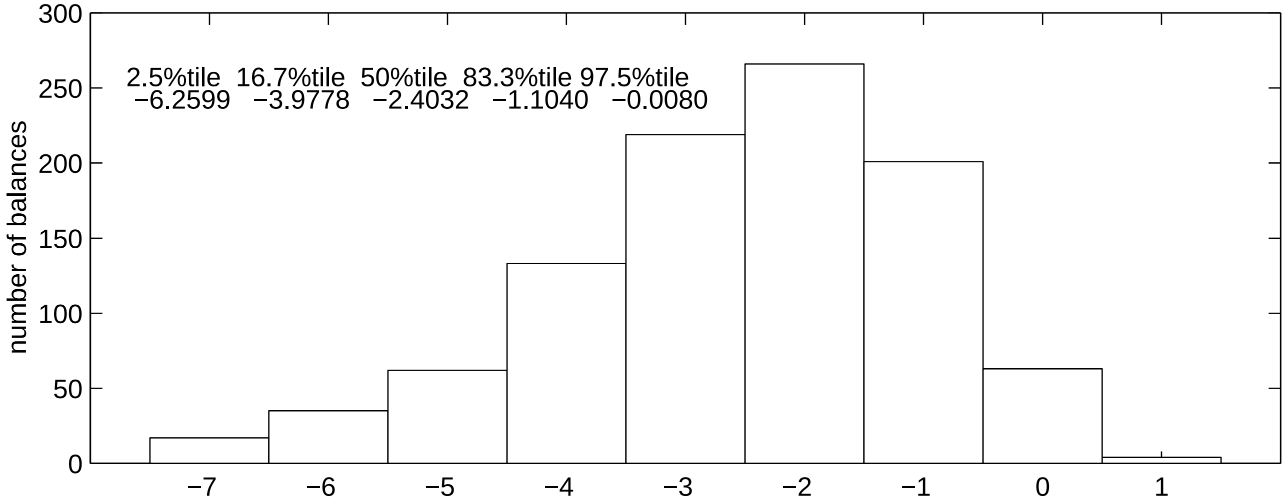


Panel B − 20% invested in S&P in 2000; 50% by 2010

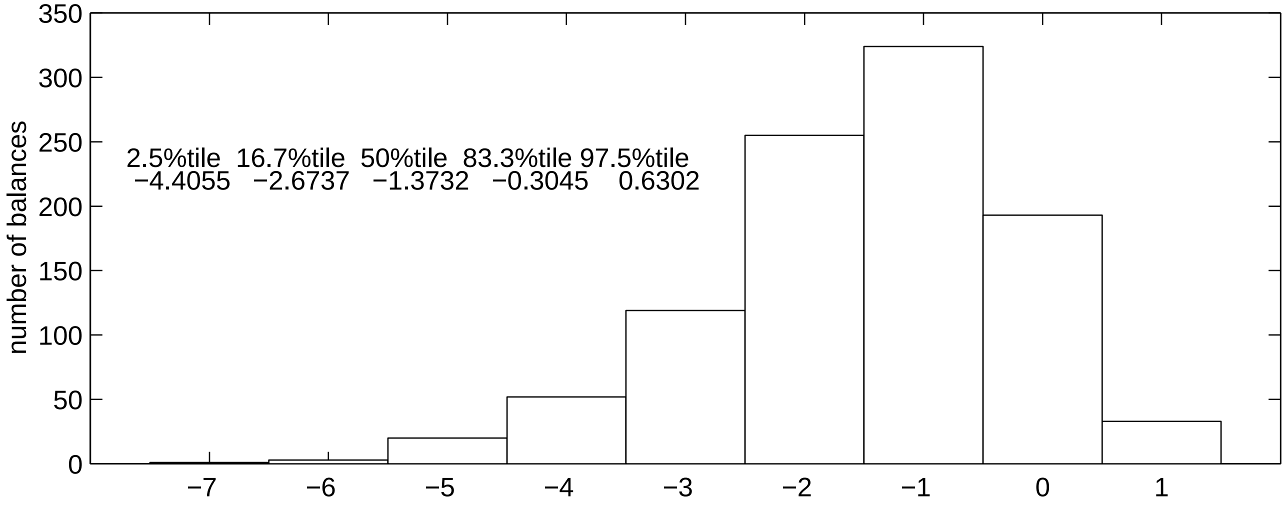


Panel C − 90% invested in S&P in 2000

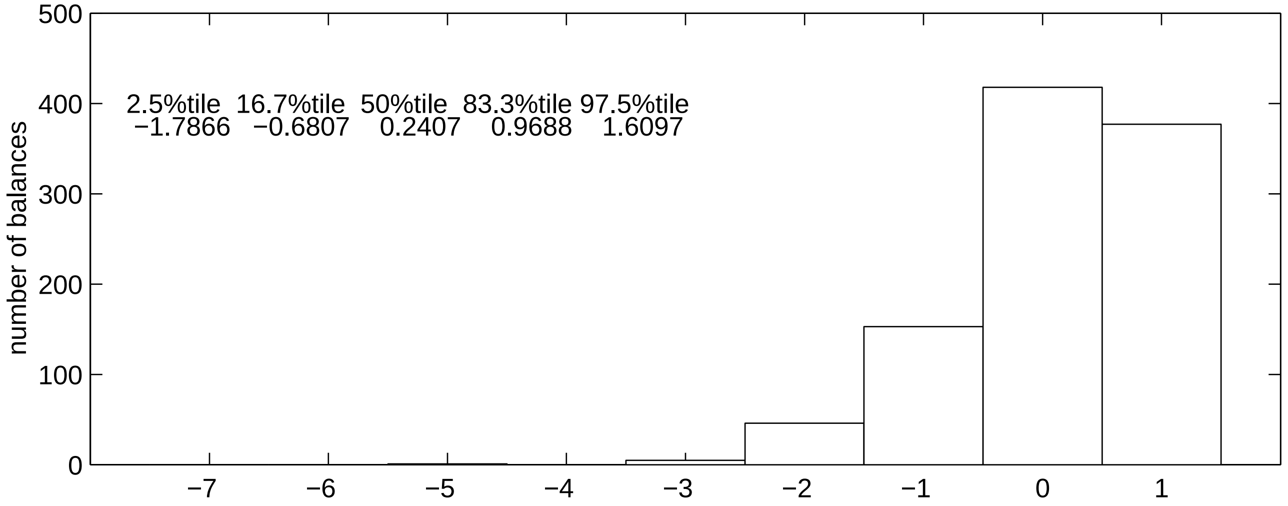
Fig. 15 − Histograms of 1,000 actuarial balances in 2072 for three scenarios



Panel A − Legislated NRA of 67 by 2022



Panel B − Accelerated NRA of 70 by 2033



Panel C − Accelerated NRA of 71 by 2022

Note: Balances expressed as percentages.