Representative Annual Survival Probabilities for Heterogenous-Agent Economies

(Revised: January 8, 2014)

The world is in a constant state of demographic flux. These demographic changes have profound effects on economic growth, fiscal policy, returns to labor and capital, and asset prices. Dynamic heterogeneous-agent models are widely used to study these issues. Demographic variables like mortality and life expectancy are an important ingredient in heterogenous agent models because they affect individual decisions, the composition and aggregation of individual decisions, and cohort distributions.

A major shortcoming of many existing studies has been that it is difficult to compare the calibration of structural parameters and elasticities in these heterogeneous-agent models to the calibrated parameters and elasticities of representative-agent frameworks. It is also hard to compare and scrutinize the quantitative results across studies because they adopt different approaches to building in demographic variables.

Mortality and life expectancy affect individuals' decision, but conditional mortalities used in current studies are specific for particular countries at particular points in time, and are often reported at five-year cohorts/intervals.

Economic data are usually reported at one-year or higher frequency. Most models and their elasticity parameters are therefore calibrated to match moments at this or higher frequency. There is no obvious way of comparing elasticity parameters and quantitative results of models with a five-year frequency with those of a one-year frequency. And even if survival probabilities were reported at an annual frequency in all countries, it would be hard to distinguish the effect of elasticity estimates from the particularities of the data.

A precise formula for mortality at all ages is, obviously, impossible. In order to analyze the effect of aging, it is, however, necessary to have a parsimonious representation of how age-specific mortality evolves with life expectancy at birth. Using the observation that the logarithm of mortality rates are almost linear in age Lee and Carter (1992) proposed a principal-components-based model, which has become the "leading statistical model of mortality [forecasting] in the demographic literature" (Deaton and Paxson, 2004). Lee and Carter developed their approach on historical U.S. mortality data, 1933-1987. However, the method is now being applied to all-cause and cause-specific mortality data from many countries and time periods (Girosi and King, 2008, p.34). It was used as a benchmark for the Census Bureau population forecasts (Hollmann, Mulder, and Kallan, 2000), two U.S. Social Security Technical Advisory Panels, recommended its use, or the use of a method consistent with it (Lee and Miller, 2001), and the United Nations Population Forecast used it (Li and Gerland, 2011).

This note proposes a transparent method to compute representative age-dependent survival probabilities as functions of life expectancy based on Lee and Carter (1992). This method makes it possible to compute representative sequences of mortality at an annual frequency given reported and projected life expectancies. This allows for a more transparent economic analysis of aging and facilitates comparisons of elasticity estimates and results across models at different levels of aggregation.

1 The Lee and Carter (1992) approach to forecast longevity

Lee and Carter (1992) suggested an approach to forecasting mortality changes for changes in longevity. Denoting the central death rate for age x in year t, $m(x,t)$, Lee and Carter (1992) fit this matrix of death rates by the specification

$$\ln[m(x,t)] = a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

for appropriately chosen sets of age-specific constants, a_x and b_x, and time-varying index k_t where

$$k_{t+1} = k_t + \theta + \epsilon_t, \tag{2}$$

This model evidently is underdetermined. k is determined only up to a linear transformation, b is determined only up to a multiplicative constant, and a is determined only up to an additive constant. Lee and Carter (1992) normalized the b_x to sum to unity and the k, to sum to 0, which implies that the a_x are simply the averages over time of the $\ln(m_{x,t})$. The model cannot be fit by ordinary regression methods, because there are no given regressors; on the right side of the equation we have only parameters to be estimated and the unknown index $k(t)$. As Lee and Carter (1992) point out, the optima can be found via a singular value decomposition (SVD) of the matrix of centered age profiles.

2 Representative mortality for economic analysis

Whereas the evolution of longevity is a first-order question for demographers, there is not much economists can add to this question. As economists we are mainly concerned with understanding the economic implications of the demographic changes the demographers are predicting.

With the life expectancies at birth that demographers report we would like to compute a transparent and representative sequence of one-year survival probabilities. The time-varying index k_t can therefore be replaced by the reported life expectancy at birth e_0.

It is well known among demographers that death rates increase exponentially with age, or, equivalently, that the logarithm of death rates increases linearly with age. It was previously assumed that mortality at advanced ages deviated from this log-linear relationship ("mortality deceleration"), but as noted by among others Keilman (1997) official statistics have systematically underpredicted number of old-age individuals. According to Gavrilov and Gavrilova (2011), as more data and better statistical methods have become available, "mortality deceleration" appear to be an artifact of mixing together several birth cohorts with different mortality levels and using cross-sectional instead of cohort data.

From various data sources, we have annual data for realized and projected life expectancy. Mortality probabilities (and accordingly survival probabilities) are much less frequent. We suggest the following regression to predict survival probabilities for years in which they are not available.

$$\log[m(x, e_0)] = \alpha + \beta^{e_0} x + \varepsilon_{x,t}, \tag{3}$$

where

$$\beta^{e_0} = \gamma + \theta e_0 + \epsilon_{e_0}, \tag{4}$$

All right-hand side variables are observable which allows ordinary regression methods to be used.

We suggest a two-stage estimation procedure. The first stage is an ordinary linear regression. In this stage, the ranking criterion for the results is out-of-sample absolute deviation between life-expectancy at birth life inputted to the model and life-expectancy at birth life predicted by the model. In the second stage, a simulated method of moments procedure provides exact estimates of life expectancy at birth.

2.1 Data

The World Health Organization Mortality Database provides data on five-year age-specific all-cause mortality and life expectancy at birth by the member states in 1990, 2000, and 2010.

Summary statistics are presented in Table 1. Average life expectancy has increased from 1990 to 2010, both measured as the average, the media, and the max across countries. The difference between the mean and the media also indicates negative skewness in the sample of countries.

Figure 1 shows death rates for Japan, which have the most data on advanced age mortality. The figure shows how death rates increase log-linearly with age for all ages after early adulthood.

2.2 First-stage estimation results

The results of estimating Equations (3) and (4) are presented in Column 1 of Table 1. All point estimates are highly significant and the R^2 is .79.

Mortality during the first year of life is substantially higher than at immediately higher ages. A dummy variable is included for mortality for the first year of life

$$\log[m(x, e_0)] = \alpha + \beta^{e_0} x + \delta_1 \mathbf{1}_{x=1} + \varepsilon_{x,t}, \tag{5}$$

The results for this specification are presented in Column 2 of Table 1. All point estimates are highly significant and the R^2 improves to .93.

As is apparent from Figure 1 and well documented in the literature, mortality rates are particularily high among teenagers and young adults and that for these age groups death rates deviate from log-linear relationship. For these age groups, a set of dummy variables are included. All point estimates are still highly significant and the R^2 improves to .96.

Though highly significant and negative, the estimated coefficient for the interaction between the slope and life expectancy at birth in the three first specifications is small. In the fourth specification, this interaction term is substituted with variables that allow the intercept and slope to be functions of life expectancy. All estimates are still highly significant and the R^2 improves further to .98.

In order to better match mortality in the first year, a quadratic term is added for infant mortality and a linear term for mortality of children between the ages of 1 and 5. The results for this specification is reported in Column 5. The R^2 increases slightly to .99.

2.3 Second-stage estimation results

Based on the predictions of the model, age-specific death rates are functions of given life expectancy. With these age-specific death, rates life tables can be constructed and implied life expectancies computed. Figure 2 shows the difference between given life expectancy e_i and predicted life expectancy $\tilde{e}(e_i)$ for the latest specification estimated in the first stage.

In the second stage, keeping all other estimates fixed, the intercept, α^{e_0}, is re-estimated with the following functional form

$$\tilde{\alpha}^{e_0} = \lambda_0 + \lambda_1 e_0 + \lambda_2 (e_0)^2 \tag{6}$$

in order to minimize deviations between given life expectancy e_i and predicted life expectancy $\tilde{e}(e_i)$

$$\min \sum_i \|\tilde{e}(e_i) - e_i\| \tag{7}$$

The objective function was minimized with the following estimates for $\tilde{\alpha}^{e_0}$

$$\tilde{\alpha}^{e_0} = -4.46 + 0.127e_0 - 0.00091(e_0)^2$$

Figure 3 shows the difference between given life expectancy e_i and predicted life expectancy $\tilde{e}(e_i)$ after the second stage estimation.

$$\log[m(x, e_0)] = -5.426 - 0.057e_0 + 0.088x + \ldots$$
$$\ldots + \begin{cases} -1.151 + 0.287e_0 - 0.003e_0^2 & \text{if } x = 1 \\ \frac{5-x}{5}\left(-1.151 + 0.287e_0 - 0.003e_0^2\right) + \frac{x}{5}\left(10.632 - 0.116e_0\right) & \text{if } 1 < x \leq 5 \\ \frac{10-x}{5}\left(10.632 - 0.116e_0\right) + \frac{x-5}{5}\left(1.002\right) & \text{if } 5 < x \leq 10 \\ \frac{15-x}{5}\left(1.002\right) + \frac{x-10}{5}\left(0.47\right) & \text{if } 10 < x \leq 15 \\ \frac{20-x}{5}\left(0.47\right) + \frac{x-15}{5}\left(0.887\right) & \text{if } 15 < x \leq 20 \\ \frac{25-x}{5}\left(0.887\right) + \frac{x-20}{5}\left(0.774\right) & \text{if } 20 < x \leq 25 \\ \frac{30-x}{5}\left(0.774\right) + \frac{x-25}{5}\left(0.443\right) & \text{if } 25 < x \leq 30 \\ \frac{35-x}{5}\left(0.443\right) & \text{if } 30 < x \leq 35 \end{cases} \tag{8}$$

The resulting representative survival probabilities at different ages for different life expectancies at birth are plotted in Figure 4.

3 Conclusion

Across the world, longevity has increased substantially. Demographers project that longevity will continue increasing in the years to come. This raises a range of important economic questions. Representative sequences of mortality rates at one-year frequency is crucial for quantitatitve economic analysis of the effects of aging and comparisons across models. This note has presented a parsimonious method for economists to compute representative annual mortality rates as function of life expectancy, taking demographers' projections of life expectancy at birth as given.

References

Deaton, A. S. and C. Paxson (2004). Mortality, income, and income inequality over time in Britain and the United States. In *Perspectives on the Economics of Aging*, NBER Chapters, pp. 247–286.

Gavrilov, L. A. and N. S. Gavrilova (2011). Mortality measurement at advanced ages: A study of the Social Security Administration death master file. *North American Actuarial Journal 15*(3).

Girosi, F. and G. King (2008). *Demographic Forecasting*. Princeton: Princeton University Press.

Hollmann, F. W., T. J. Mulder, and J. E. Kallan (2000). Methodology and assumptions for the population projections of the United States: 1999 to 2100. U.S. Census Bureau Population Division Working Paper 38.

Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics 13*(3), 245–277.

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association 87*.

Lee, R. D. and T. Miller (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography 38*(4), 537–549.

Li, N. and P. Gerland (2011). Modifying the Lee-Carter method to project mortality changes up to 2100. *Annual Meeting of the Population Association of America*.

Table 1: Summary statistics

	1990	2000	2010
mean	71.60	73.75	75.07
median	73.20	76.05	78.00
st.dev.	5.92	6.10	6.83
min	54.5	58.2	51.3
max	79.1	81.3	82.6
n obs.	44	44	44

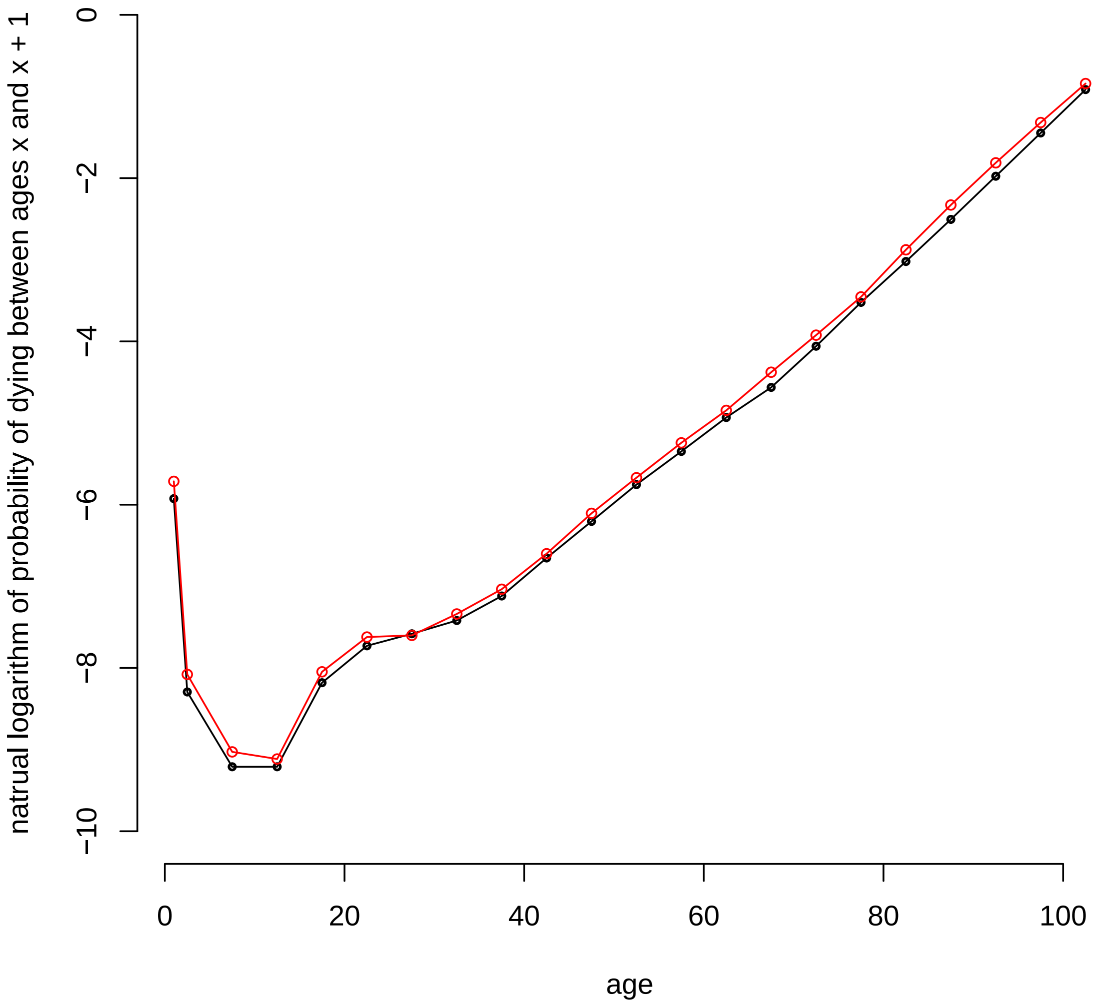

Figure 1: Natural logarithm of 1 - death rates: Japan

Table 2: First-stage regression results

	1 Estimate (S.E.)	2 Estimate (S.E.)	3 Estimate (S.E.)	4 Estimate (S.E.)	5 Estimate (S.E.)
(Intercept)	-7.884*	-8.594*	-9.534*	-5.032*	-5.426*
	(0.064)	(0.039)	(0.058)	(0.129)	(0.116)
γ	0.11*	0.12*	0.133*	0.088*	0.088*
	(0.007)	(0.004)	(0.003)	(0.001)	(0)
θ	-0.001*	-0.001*	-0.001*	.	.
	(0)	(0)	(0)		
δ_0	.	4.437*	5.327*	11.064*	-1.151
		(0.098)	(0.089)	(0.578)	(4.724)
δ_1	.	.	2.361*	2.361*	10.632*
			(0.089)	(0.059)	(0.511)
δ_2	.	.	1.002*	1.002*	1.002*
			(0.086)	(0.058)	(0.051)
δ_3	.	.	0.47*	0.47*	0.47*
			(0.084)	(0.056)	(0.05)
δ_4	.	.	0.887*	0.887*	0.887*
			(0.082)	(0.055)	(0.048)
δ_5	.	.	0.774*	0.774*	0.774*
			(0.08)	(0.054)	(0.047)
δ_6	.	.	0.443*	0.443*	0.443*
			(0.078)	(0.052)	(0.046)
α^{e_o}	.	.	.	-0.063*	-0.057*
				(0.002)	(0.002)
$\delta_0^{e_o}$	.	.	.	-0.08*	0.287*
				(0.008)	(0.139)
$\delta_1^{e_o}$	.	.	.	.	-0.116*
					(0.007)
$\delta_0^{(e_o)^2}$	.	.	.	.	-0.003*
					(0.001)
N	968	968	968	968	968
$RMSE$	1.054	0.597	0.455	0.305	0.269
R^2	0.791	0.933	0.961	0.983	0.987
adj R^2	0.791	0.933	0.961	0.983	0.986

* $p \leq 0.05$

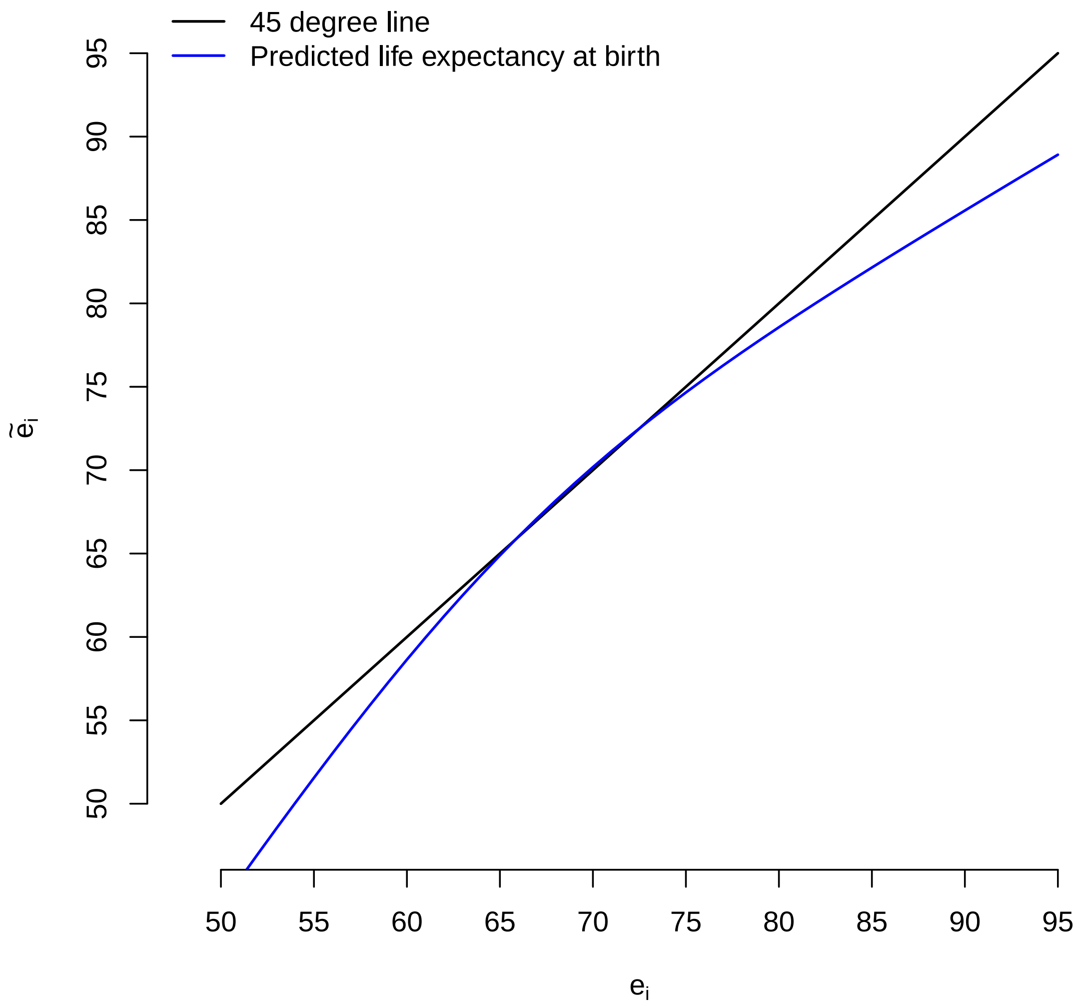

Figure 2: After first-stage estimation: given and predicted life expectancy at birth

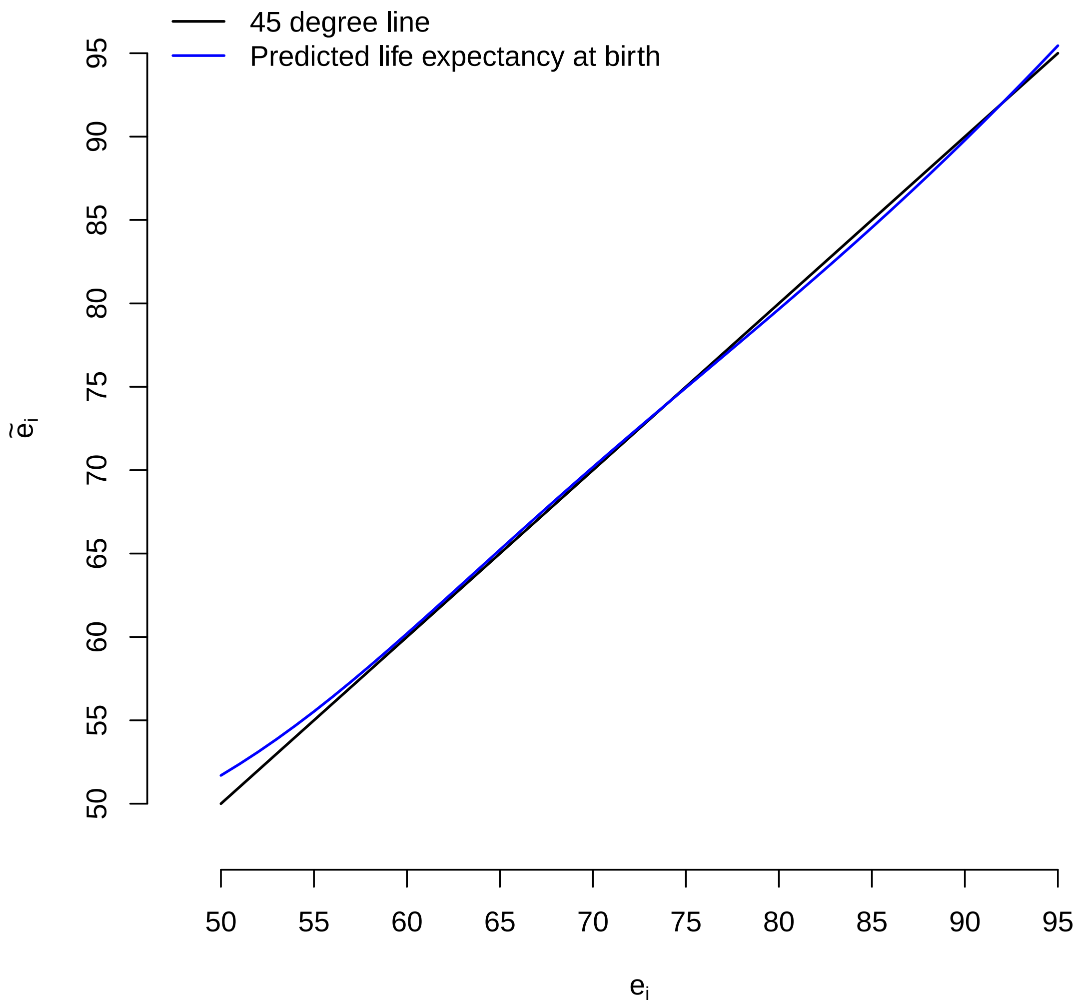

Figure 3: After second-stage estimation: given and predicted life expectancy at birth

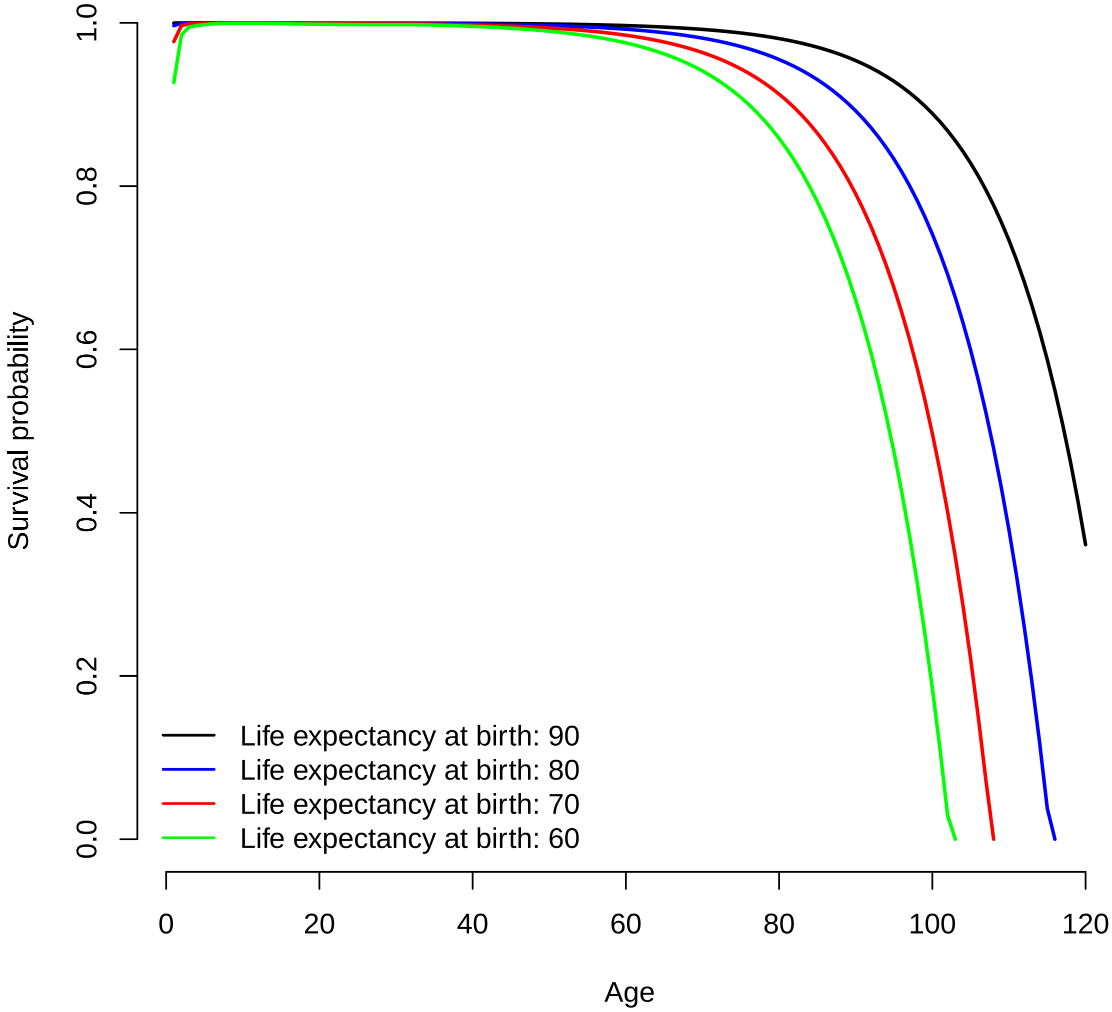

Figure 4: After second-stage estimation: given and predicted life expectancy at birth