GENUS, **LXV** (No. 1), 23-57

FRANCO PERACCHI* – VALERIA PEROTTI**

Subjective survival probabilities and life tables: an empirical analysis of cohort effects

1. INTRODUCTION

Survival probabilities are a crucial ingredient in the analysis of economic issues that involve life-cycle decisions under uncertainty. Longevity can be highly heterogeneous within a population, depending on both observed and unobserved characteristics. Part of this variability can be explained in the framework proposed by Ehrlich (2000) and Benítez-Silva and Ni (2008), in which longevity is regarded as the output of a production function with biological initial conditions, health-related behavior, and health investments as inputs (Grossman, 1972).

Several studies have been carried out to assess how longevity changes with individual characteristics such as education, income, employment status, and health status. Because this detailed information is usually not available from life tables, that only report average survival probabilities at the population level or at the level of a few population subgroups, researchers have tried to estimate survival probabilities from longitudinal sample surveys. Although these data provide estimates that are subject to sampling variability and model uncertainty, they offer the opportunity of controlling for a very rich set of socio-demographic and economic characteristics. As noted by Delavande and Rohwedder (2008), however, this approach requires large samples of individuals who are interviewed over a long period of time with low attrition rates.

An attractive alternative is to use subjective survival probabilities, which are now asked in several household surveys. An additional reason for using subjective survival probabilities instead of life tables is that economic decisions are likely to be based on individual beliefs, and these beliefs may be based on more accurate information on own longevity than the life tables (Perozek, 2008). Previous work has shown that subjective survival probabilities are systematically related to a number of risk factors (Hurd and McGarry, 1995 and 2002; Benítez-Silva and Ni, 2008), are updated by individuals in response to health shocks or new information (Smith *et al.*,

*University of Rome " Tor Vergata" and Einaudi Institute for Economics and Finance (EIEF)

**Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori (ISFOL)

2001; Hurd and McGarry, 2002; Benítez-Silva and Ni, 2008), and help predict actual mortality (Smith *et al.*, 2001; Hurd and McGarry, 2002; Elder, 2007). Delavande and Rohwedder (2008) compare a model for survival to age 75 based on actual mortality and one for subjective probabilities of survival to age 75 estimated for the same sample of individuals, and find similar coefficients on wealth, income and education indicators.

An important question is whether, on average, subjective probabilities are consistent with life tables. An early attempt to measure subjective survival probabilities and to compare them to actuarial data was Hamermesh (1985), who carried out a small survey of expected age at death and survival probabilities among economists. More recent studies systematically compare subjective assessments from household surveys and life-table probabilities. For example, Hurd and McGarry (1995, 2002) and Elder (2007) use data from several waves of the US Health and Retirement Study (HRS), whereas Guiso *et al.* (2005) and Balia (2007) extract information from the 2004 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE).

Common findings are that subjective probabilities of survival to age 75 tend to be lower than life-table probabilities, whereas the opposite is true for survival to older ages. Subjective probabilities also tend to be higher than life-table probabilities for males, whereas the opposite is true for females. Survival probabilities also tend to be higher for individuals with better self-assessed health status, for people with higher educational attainments, with parents who are still alive or died at older ages, for people who are physically active, do not smoke, and drink less (Hurd and McGarry, 1995, 2002), and for those who are either retired or employed (Balia, 2007). Using the 2006 wave of the HRS, that asks respondents to assess the survival probability of people of their same age and gender, Elder (2007) finds that subjective assessments are systematically different from the 2003 life-table values, particularly at younger ages. He interprets these findings as evidence of a bias in subjective survival probabilities.

The main drawback of all these studies is that they compare subjective survival probabilities to data from either a single life table or a set of life tables, none of which may be taken to represent the survival profile of a particular cohort. As remarked by Elder (2007), “life tables at a point in time do not represent the actual mortality profile facing a particular cohort if age-specific mortality rates decline over time”, as it is the case for the US and most developed countries. This is also true for time averages of life-table data. An exception in the literature is Perozek (2008), who constructs survival functions based on subjective probabilities from the HRS and compares them to data from cohort life tables. Her main result is that male subjective probabilities are roughly in line with the life tables, whereas female subjective probabilities are lower than those reported in the life tables. However, cohort life tables are based on mortality forecasts which may not be sufficiently accurate.

The purpose of this paper is to study both potential bias and heterogeneity in subjective survival probabilities. Our first step is to analyze the behavior of average subjective survival probabilities relative to survival probabilities from life tables. In our second step, due to the lack of finely disaggregated life-table data, we can only focus on heterogeneity in subjective survival probabilities. Our data on subjective probabilities come from the 2006 wave of the Italian Participation, Labour and Unemployment Survey (PLUS), a large representative survey of the Italian working-age population. Given the absence of cohort life tables for Italy, we construct cohort-specific survival probabilities using a set of life tables spanning a relatively long time period (1982-2003), thus avoiding the problems that plague comparisons which do not control for the presence of cohort effects in survival. After showing that subjective probabilities are on average broadly consistent with life tables, we study how they vary with observed individual characteristics, such as education, marital status, health status, *etc*., using a reduced-form framework similar to Hurd and McGarry (1995) and Benítez-Silva and Ni (2008), but adopting a discrete-choice approach and accounting for potential sample selection bias due to item nonresponse.

The rest of the paper is organized as follows. Section 2 describes our life-table and subjective survival probability data. Section 3 presents the method that we use to construct cohort-specific probabilities from life tables. Section 4 compares average subjective survival probabilities to predictions from cohort life tables, and then analyzes heterogeneity in subjective survival probabilities. Finally, Section 5 offers a summary and some concluding remarks.

2. DATA

2.1 The ISTAT life tables

The Italian National Institute of Statistics (ISTAT) publishes annual cross-sectional life tables for the Italian population, broken down by single year of age, gender and broadly-defined geographical area (Northeast, Northwest, Center, and South/Islands). The set of life tables that we use in this paper spans the period from 1982 to 2003. For each age x from 0 to 120 years, these tables report the probability of dying at age x (conditional on surviving until that age) and residual life expectancy. In addition, the table compiled for a given year t reports the survival profile of a hypothetical cohort of individuals, whose mortality at a given age corresponds to the actual mortality of Italians of the same age as measured in year *t*. As we discuss in Section 3, this hypothetical cohort does not correspond to any particular cohort and, therefore, the cross-sectional survival profile cannot be used as benchmark for evaluating subjective survival probabilities.

Since ISTAT only publishes cross-sectional life tables and does not provide cohort life tables for cohorts that are still alive, we need to predict the future mortality of these cohorts. We do this by exploiting the information on their past mortality, obtained by taking advantage of the long time-series dimension of the available set of cross-sectional life tables. Section 3 discusses alternative methods for constructing these longitudinal predictions.

2.2 The PLUS data

Subjective survival probabilities are obtained from the microdata of the second (2006) wave of the Participation, Labour and Unemployment Survey (PLUS), a cross-sectional survey with a longitudinal component, broadly representative of the Italian working-age population[1] carried out by ISFOL (Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori) in 2005 and 2006. In addition to standard socio-demographic information (demographics, labor force status, education and training, earnings, *etc*.), the 2006 wave of PLUS asks respondents to report their subjective assessment of their own probability of reaching age 75 and age 90, on a 0-100 scale.

The PLUS sample contains 37,513 individuals aged 15-64 (16,825 men and 20,688 women). Because ISFOL planned to re-interview part of the sample in subsequent waves, interviewers were instructed not to ask the questions on survival probabilities in cases where this might have been perceived as indelicate or inappropriate. As a result, no question was asked to 11.0 percent of the sample, only one question was asked to 0.3 percent, and both questions (survival to age 75 and to age 90) were asked to 88.6 percent. Of those who were asked both questions, 22.8 percent answered none, 3.1 percent answered only one, and 74.1 percent answered both. We exclude from the sample the small fraction (less than 1 percent) of respondents with inconsistent answers, namely those reporting a probability of survival to age 75 lower than the probability of survival to age 90. Overall, we have 24,466 individuals (10,921 men and 13,545 women) providing consistent subjective survival probabilities. Since we want to analyze the relationship between survival variability and a set of individual characteristics, we further drop individuals for whom these characteristics

[1] People aged 15-64, with the exception of a few categories, such as inactive women aged 40-64 who were not retired. A more detailed description of the survey can be found in Appendix A.

are not available due to nonresponse (3.3 percent).

Our final sample consists of 23,657 individuals (10,486 men and 13,171 women), of whom 10,177 are currently working. Table 1 presents summary statistics for the set of covariates that we use. This set includes age, gender,

Table 1 – *Summary statistics for the covariates in the regression models*

Variable	No. Obs.	Mean	Std. Err.	Min	Max
Age	23657	35.45	14.56	15	64
Female	23657	0.56	0.50	0	1
Region					
North	23657	0.42	0.49	0	1
Center	23657	0.19	0.39	0	1
South	23657	0.39	0.49	0	1
Education					
Primary or less	23657	0.28	0.45	0	1
Secondary	23657	0.53	0.50	0	1
Tertiary	23657	0.19	0.40	0	1
Marital status					
Married/cohabitant	23657	0.44	0.50	0	1
Not married	23657	0.56	0.50	0	1
Self-reported diseases					
None	23657	0.98	0.14	0	1
Short term	23657	0.01	0.11	0	1
Long term	23657	0.01	0.09	0	1
Activity status					
Employed	23657	0.43	0.50	0	1
Unemployed	23657	0.15	0.36	0	1
Inactive	23657	0.42	0.49	0	1
Employment status					
Long term employee	10177	0.64	0.48	0	1
Short term employee	10177	0.24	0.43	0	1
Self-employed	10177	0.12	0.33	0	1
Gross annual earnings	10177	21698.86	17964.76	300	375000
Risky job	10177	0.27	0.44	0	1
Satisfied with workload	10177	0.77	0.42	0	1
Satisfied with safety	10177	0.79	0.41	0	1
Satisfied with earnings	10177	0.51	0.50	0	1
Satisfied with stability	10177	0.77	0.42	0	1

geographic region of residence (North, Center, South), education (primary or less, secondary, tertiary education), marital status (married/cohabitant, not married), self-reported diseases (none, short term, long term), and activity status (employed, unemployed, inactive). For individuals who are currently employed, the set of covariates also includes the logarithm of gross annual earnings and indicators for employment status (long term employee, short term employee, self-employed), for holding a risky job and for being satisfied with a number of job characteristics (workload, safety, earnings and job stability). In addition to nonresponse to questions on subjective survival probabilities, one drawback of our data is that, for females, comparability between subjective and life-table probabilities is limited by the fact that inactive women aged 40-64 who were not retired from the labor force were not included in the reference population of the PLUS survey (see Appendix A).

2.3 Subjective survival probabilities

Table 2 presents the mean and the 25^{th} ($Q25$), 50^{th} ($Q50$) and 75^{th} ($Q75$) percentiles of the empirical distribution of subjective probabilities of survival to age 75 and to age 90, separately by gender and age group (15-29, 30-39, 40-49, and 50-64 years). The average subjective probability of survival to age 75 is 80.2 percent (81.3 for men and 79.3 for women), whereas the average subjective probability of survival to age 90 is 52.4 percent (53.1 for men and 51.9 for women). For both men and women, the age-profile of the subjective probability of survival to age 75 is U-shaped, with a minimum in the 30-50 age range, whereas the age-profile of the subjective probability of survival to age 90 is monotonically declining. Subjective survival probabilities for women do not exceed those for men, in contrast with women's higher longevity.

Figure 1 shows the distribution of the answers to the survival probability questions. Although these answers may take any value between 0 and 100, most respondents use multiples of 5 or 10. This finding may partly be explained by rounding, but there are also focal responses at 0, 50, and 100. As shown in the figure, "0" answers are much less frequent for survival to age 75 than for survival to age 90, while the opposite is true for "100" answers. This suggests that the main explanation for these values is rounding. On the other hand, the fraction of individuals answering "50" is higher in the case of survival to 90 than in the case of survival to 75, and "50" is also the most frequent answer in the former case. Bruine de Bruin *et al.* (2002) argue that answering "50" to a subjective probability question may reflect epistemic uncertainty rather than a probabilistic belief, especially if the question is open-ended. A similar result is found by Lillard and Willis (2001), who assume that respondents are uncertain about the true probability and provide the most likely among all possible values (modal response hypothesis, MRH). Using data from the HRS, Hill *et al.* (2004) present maximum likelihood estimates of a parametric model for the subjective survival probability based on the MRH, and find that higher education and higher cognitive scores are associated with lower uncertainty.

Table 2 – *Mean and percentiles of subjective survival probabilities by gender and age group*

Age group	Age 75				Age 90			
	Mean	*Q*25	*Q*50	*Q*75	Mean	*Q*25	*Q*50	*Q*75
				Men				
15-29	81.4	70	85	100	55.9	30	60	80
30-39	80.1	70	85	100	52.5	30	50	80
40-49	79.6	70	80	100	50.8	20	50	80
50-64	81.9	70	90	100	50.1	20	50	80
Total	81.3	70	90	100	53.1	30	50	80
				Women				
15-29	80.4	70	80	100	55.6	40	50	80
30-39	76.5	60	80	100	49.0	20	50	70
40-49	78.9	70	80	100	48.7	20	50	80
50-64	80.9	70	90	100	47.9	20	50	80
Total	79.3	70	80	100	51.9	30	50	80
				Total				
15-29	80.8	70	80	100	55.7	35	55	80
30-39	77.4	60	80	100	49.8	20	50	75
40-49	79.2	70	80	100	49.6	20	50	80
50-64	81.6	70	90	100	49.3	20	50	80
Total	80.2	70	80	100	52.4	30	50	80

After estimating three logistic regression models for the probability of answering “0”, “50” and “100”, we conclude that the “0” and “100” answers can be explained by rounding, as they are systematically related to health status and other variables that are expected to affect beliefs about own survival[2]. On the other hand, at least part of the “50” answers may be due to epistemic uncertainty.

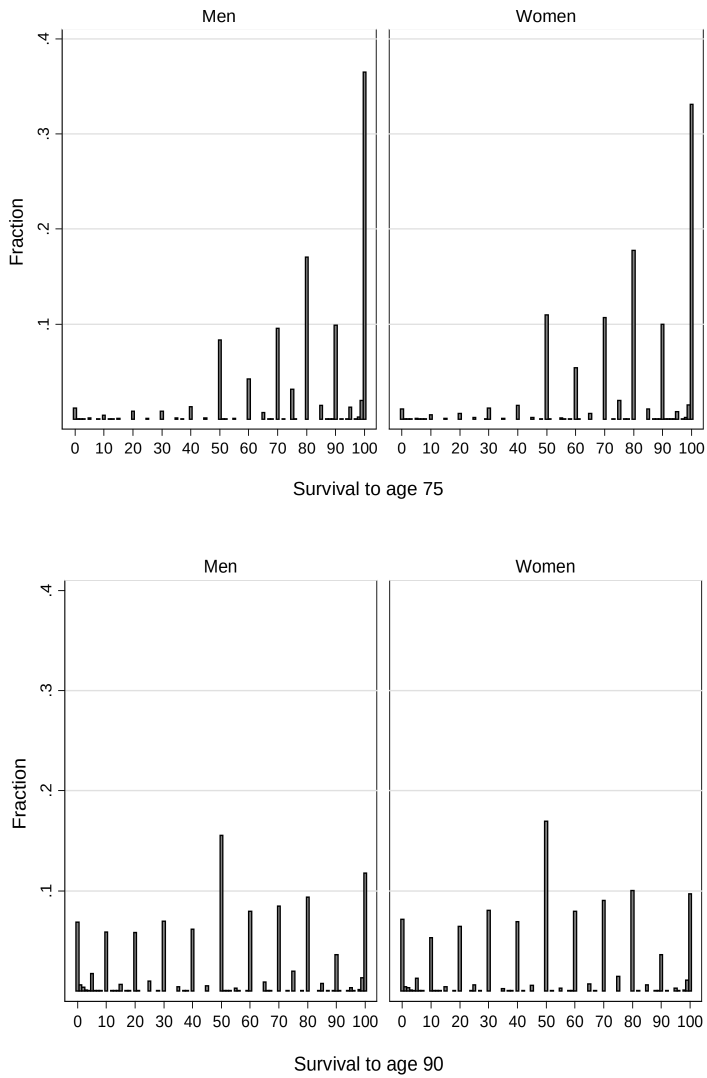
3. AVERAGE SUBJECTIVE SURVIVAL PROBABILITIES AND LIFE TABLES

If the proper set of survey weights is used, then the sample average of subjective survival probabilities from a cross-sectional survey such as PLUS would provide an unbiased estimator of the mean of subjective survival probabilities in a population. Mean subjective survival probabilities, however, need not coincide with the probabilities of future survival derived from a cohort life table. This may occur either because individuals form biased expectations of their own survival, or because they have access to

[2] Results are available upon request.

more accurate information about future survival than what is used to construct the cohort life tables (Perozek, 2008).

Figure 1 – *Subjective survival probabilities in the PLUS sample*



In the case of Italy, a comparison of subjective and objective survival probabilities is complicated by the fact that cohort life tables are not available for cohorts that are still alive. So, an important preliminary issue is how to use the available set of cross-sectional life tables to construct the probabilities of future survival for these cohorts.

Life tables provide information on q_x, the probability of dying before reaching age $x+1$ conditional on survival up to age *x*. If we had access to a cohort life table, that is, to the entire profile of q_x for a given cohort, then we could compute the cohort-specific conditional probability at age *a* of surviving to age $x+1$ as

$$\Pr\{M > x|M > a-1\} = \frac{\Pr\{M > x\}}{\Pr\{M > a-1\}} = \prod_{i=a}^{x}(1-q_i),$$

where *M* is a random variable representing age at death. Suppose now that we only have access to a single cross-sectional life table compiled for year *t*. Because the life table provides estimates of $q_x = D_x^t/N_x^t$, where N_x^t is the number of people who reach age *x* in year *t* and D_x^t is the number of people who do not reach age $x+1$, the product of survival probabilities from the cross-sectional life table is

$$P_C = \left(1-\frac{D_x^t}{N_x^t}\right)\left(1-\frac{D_{x-1}^t}{N_{x-1}^t}\right)\cdots\left(1-\frac{D_a^t}{N_a^t}\right) = \prod_{i=a}^{x}(1-q_i^{t-i}), \qquad [1]$$

where the superscript $t-i$ indexes the year of birth. The right-hand side of [1] is the product of survival probabilities of different cohorts and so P_C does not generally represent the survival probability of any particular cohort. In what follows, we refer to P_C as the *cross-sectional* prediction of survival. If the variable *B* denotes the year of birth of a cohort, then the probability of surviving to age $x+1$ for people born in year $B=b$ who survived to age *a* is instead

$$P_L = \Pr\{M > x|M > a-1, B=b\} = \prod_{i=a}^{x}(1-q_i^b). \qquad [2]$$

We refer to P_L as the *longitudinal* prediction of survival. In practice, this prediction cannot be constructed because we do not know the cohort-specific death probabilities at ages beyond the current age.

A variety of approaches may be used to recover these cohort-specific death probabilities from a sufficiently long sequence of cross-sectional life tables. One approach consists of estimating a model with a set of cohort dummies among the covariates. Since each cohort is observed over a different age range, the assumption of fixed cohort effects may create problems when there are age-specific changes in mortality. One such example is the trend towards increased mortality of young Italian men and women observed between the mid 1980’s and the mid 1990’s (bottom panel of Figure 2). A model with cohort dummies would spread this age-specific increase in mortality to future ages.

A second approach consists of fitting an *r*-th order polynomial trend to the log-odds of dying at age *i*, separately for each age *i*,

$$\log \frac{q_i^{t-i}}{1-q_i^{t-i}} = \beta_{0i} + \beta_{1i} t + \beta_{2i} t^2 + \cdots + \beta_{ri} t^r + \varepsilon_{it}. \qquad [3]$$

This model captures the trend in mortality across cohorts, but ignores cohort-specific effects.

A third approach, proposed by Lee and Carter (1992), consists of several steps. First, the logarithm of death probabilities is modelled as a linear function of a time-varying “mortality index” χ_t, with age-specific paramaters β_{0i} and β_{1i}

$$\log q_i^{t-i} = \beta_{0i} + \beta_{1i} \chi_t + \varepsilon_{it}. \qquad [4]$$

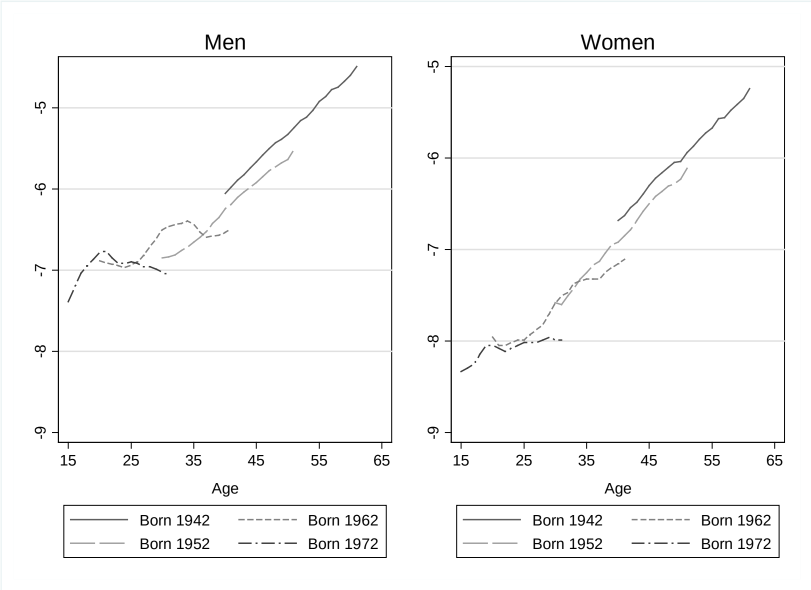
Since all the right-hand side terms are unknown, β_{0i}, β_{1i} and χ_t are estimated using a method based on singular value decomposition. Then, an ARIMA model is estimated for χ_t to predict its future values. Finally, predictions for the age-specific death probabilities are computed based on model [4].

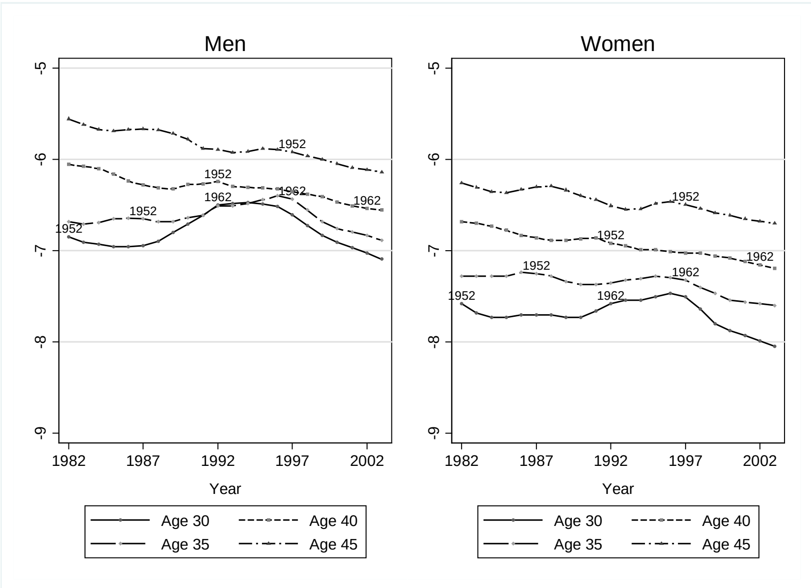
In this paper we present results based on model [3] and compare them with the results from the Lee-Carter procedure. In fact, for $r=1$, model [3] is a special case of model [4] when $\chi_t = t$ and log-odds are used instead of logarithms[3]. Denoting by $\hat{y}_i^b$ the predicted log-odds of death at age *i* for people born in year *b*, obtained from model [3], the estimated probability at age *i* of dying before age $i+1$ is

$$\hat{q}_i^b = \frac{\exp \hat{y}_i^b}{1+\exp \hat{y}_i^b}.$$

[3] When q is small, the log-odds can be approximated as $\log q + q$. In the life tables, $q < 0.075$ for all ages 0–75.

Figure 2 – *Age profiles of the log-odds of death by cohort (1982-2003 life tables)*





For a cohort born in year $B = b$, we evaluate the probability of surviving to age 75 and 90 for a respondent of age *a* by using equation [2] with q_i^b replaced by its predicted value $\hat{q}_i^b$ and *x* equal to either 75 or 90

$$\hat{P}_L = \prod_{i=a}^{x}(1-\hat{q}_i^b).$$

We refer to $\hat{P}_L$ as the *estimated* longitudinal prediction of survival. Finally, we compare the estimated longitudinal predictions with the average subjective probabilities estimated from the 2006 PLUS survey. Since these data represent a single cross-section, for each age *a* we have the subjective survival probability of just one cohort, namely the cohort born in year 2006 - α Therefore, we compare the average subjective probability provided by respondents of age *a* to $\hat{P}_L$ with $b = 2006 - a$.

4. EMPIRICAL RESULTS

In this section, we compare the average subjective probabilities provided by the PLUS respondents to the estimated longitudinal predictions of survival constructed using the method described in Section 3. We then study the relationship between heterogeneity in subjective survival probabilities and observed individual characteristics.

4.1 Predicted vs subjective survival probabilities

Estimating model [3] requires choosing the order of the polynomial trend. In our case, a linear trend is perfectly adequate for most ages. The main exception is for the 20-35 age group, due to the temporary increase and the subsequent decline of mortality between the mid 1980's and the mid 1990's for this age group. We use a linear trend model also for the 20-35 age group because fitting a quadratic model generates a decline of mortality in later years that is too steep.

Table 3 presents the estimates for selected ages using data at the national level. For the reasons we just discussed, the overall fit of the model is rather poor for ages 20 to 35, but is quite good for all other ages with very high regression R^2's. For men, the coefficient on the linear trend is highest for ages between 50 and 65, whereas for women it is highest for ages between 65 and 80. Figure 3 shows predicted death probabilities and life expectancy for people born in 1942, 1966 and 1991 who, in 2006, are aged

64, 40 and 15 respectively. These predictions are very close to those obtained by the Lee-Carter procedure with an ARIMA(0,1,0) model estimated in the second step (Figure 4)[4].

Table 3 – *Linear trend models for the log-odds of death, by gender and selected ages, at the national level*

Age	β_0	β_1	No. obs.	R^2
		Men		
15	46.279 ***	-0.027 ***	22	0.787
20	7.136	-0.007 ***	22	0.317
25	-2.171	-0.002	22	0.016
30	-10.155	0.002	22	0.003
35	-2.366	-0.002	22	0.011
40	32.833 ***	-0.020 ***	22	0.873
45	45.575 ***	-0.026 ***	22	0.950
50	55.435 ***	-0.031 ***	22	0.966
55	61.713 ***	-0.033 ***	22	0.983
60	59.205 ***	-0.032 ***	22	0.982
65	54.056 ***	-0.029 ***	22	0.964
70	45.124 ***	-0.024 ***	22	0.938
75	41.235 ***	-0.022 ***	22	0.932
80	38.724 ***	-0.021 ***	22	0.873
85	27.196 ***	-0.015 ***	22	0.829
90	28.637 ***	-0.015 ***	22	0.893
		Women		
15	24.653 ***	-0.017 ***	22	0.716
20	11.228 ***	-0.010 ***	22	0.520
25	15.824 **	-0.012 ***	22	0.383
30	10.884	-0.009 *	22	0.151
35	21.223 ***	-0.014 ***	22	0.669
40	36.337 ***	-0.022 ***	22	0.968
45	32.286 ***	-0.019 ***	22	0.884
50	32.620 ***	-0.019 ***	22	0.969
55	37.572 ***	-0.022 ***	22	0.963
60	42.634 ***	-0.024 ***	22	0.980
65	48.266 ***	-0.027 ***	22	0.980
70	50.532 ***	-0.027 ***	22	0.957
75	48.872 ***	-0.026 ***	22	0.957
80	49.124 ***	-0.026 ***	22	0.928
85	41.839 ***	-0.022 ***	22	0.899
90	43.784 ***	-0.023 ***	22	0.937

Notes:* denotes asymptotic *p*-values between 5 and 10 percent, ** denotes asymptotic *p*-values between 1 and 5 percent, *** denotes asymptotic *p*-values below 1 percent.

[4] The first step of the Lee-Carter procedure was estimated using the leecart command in Stata (Wang, 2000). Results are available upon request.

We now compare subjective survival probabilities, averaged by age, to the estimated longitudinal predictions constructed from the life tables. The analysis is carried out separately for men and for women, first at the national level and then by region.

Figure 3 – *Predicted death probabilities and life expectancy for selected cohorts in the PLUS sample (procedure based on the linear trend model)*

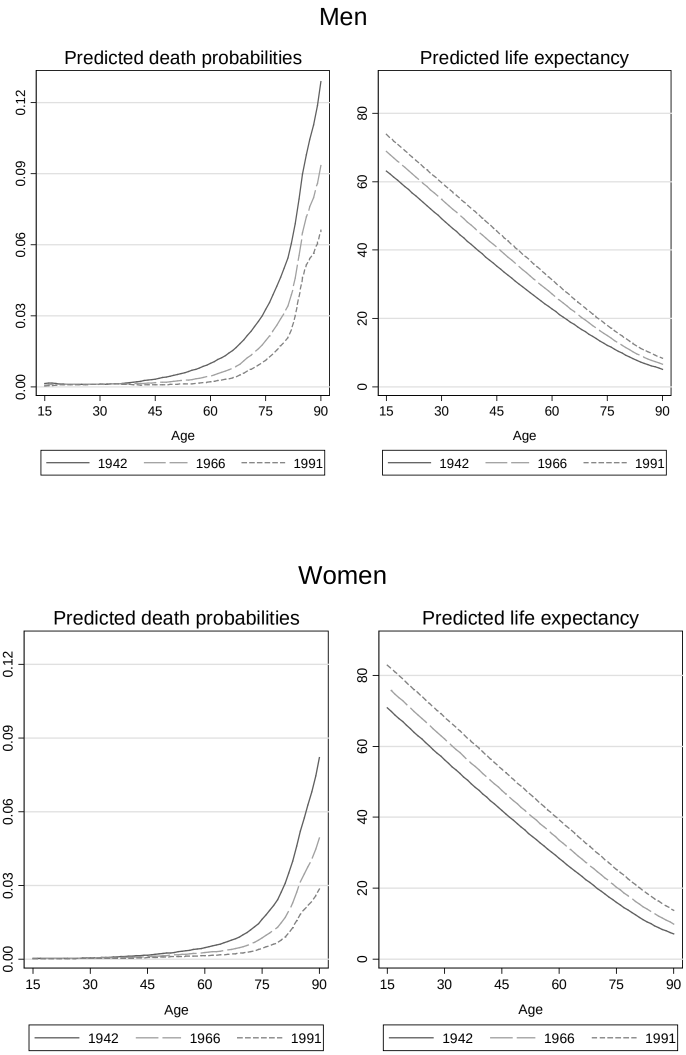
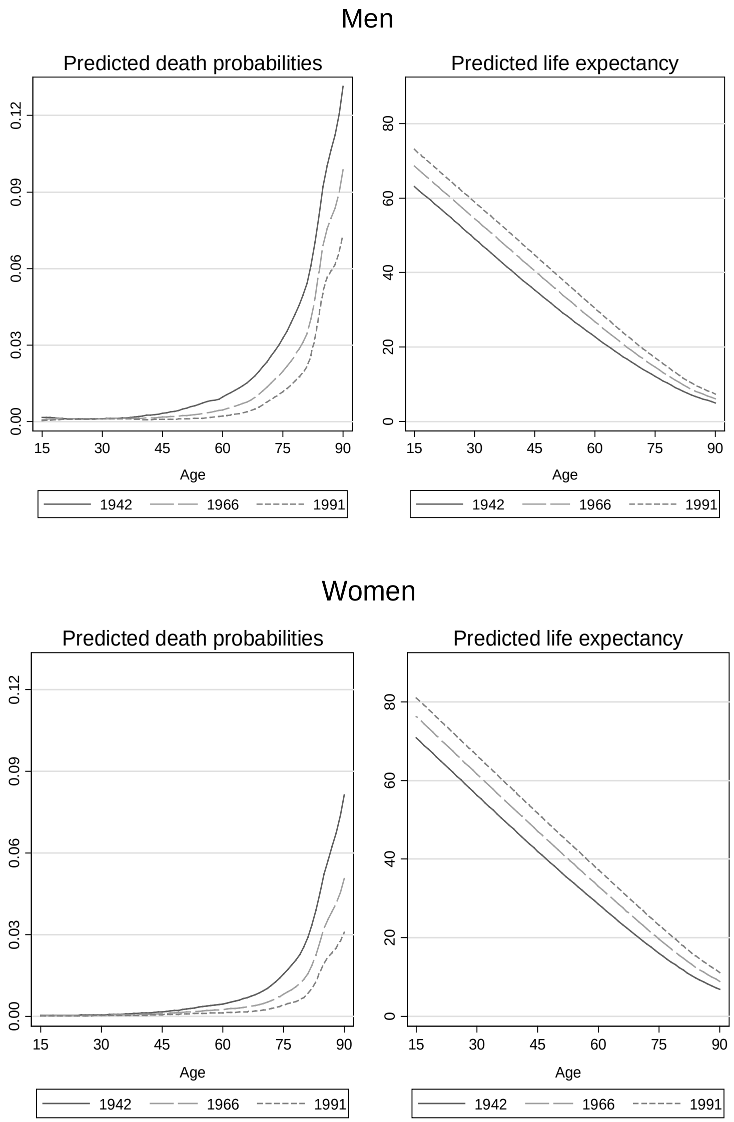


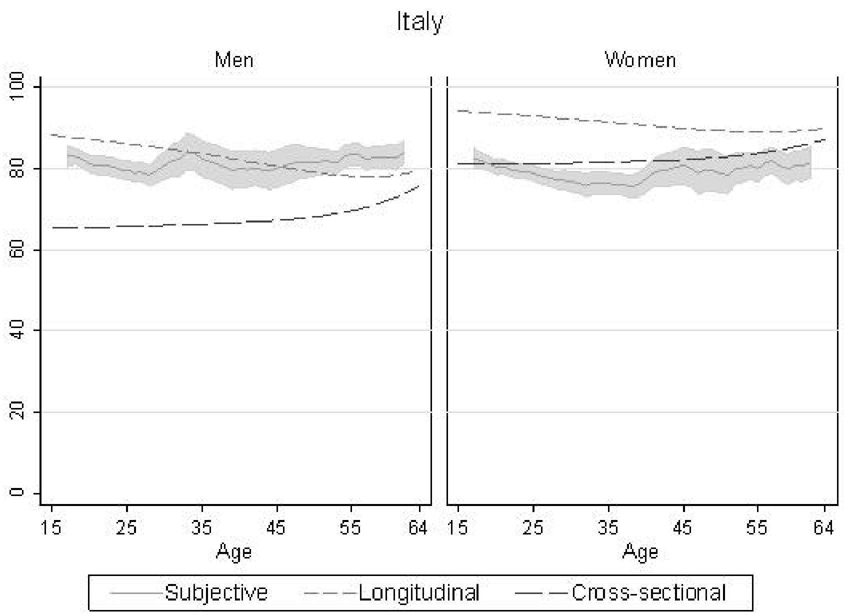
Figure 4 – *Predicted death probabilities and life expectancy for selected cohorts in the PLUS sample (Lee-Carter procedure)*



To improve comparability between the two sets of probabilities, we use the largest sample available, i.e. all individuals who provide reliable survival probabilities (24,466 respondents), without dropping people who did not answer other questions. The survey weights provided in PLUS are used throughout.

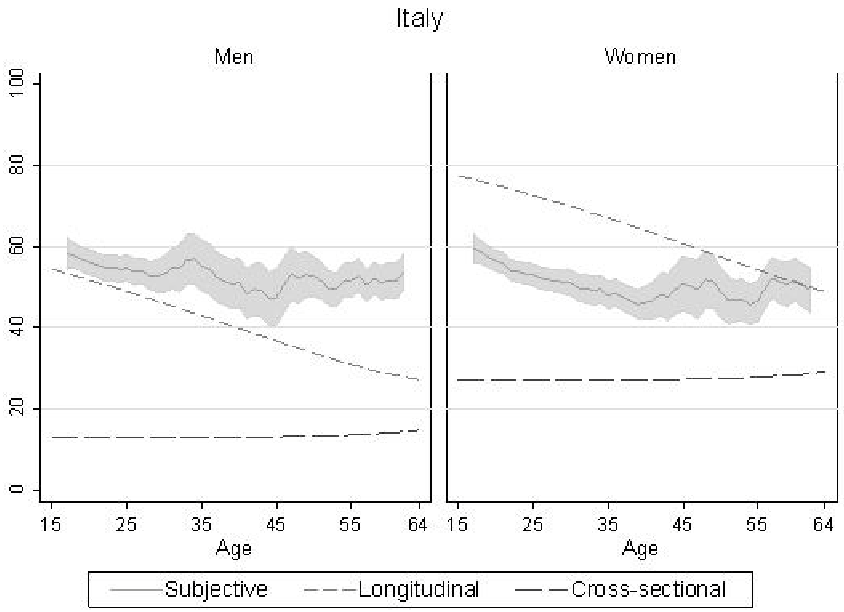
Figures 5 to 8 show a 5-year moving average of the subjective probabilities of survival (with 2-standard error bands), along with the cross-sectional predictions from the 2003 life table and the longitudinal predictions from the sequence of life tables for the years 1982-2003. If our prediction model is accurate, then the last set of predictions is the proper counterpart of subjective survival probabilities elicited in the PLUS survey. Figures 5 and 6 are for $x = 74$ (survival to age 75) and $x = 89$ (survival to age 90) respectively, with data at the national level. It should be noted that the cross-sectional and the longitudinal predictions are very different. In particular, the latter are always higher than the former, and the difference is largest at younger ages. Irrespective of gender or target age, longitudinal predictions are decreasing with age, whereas cross-sectional predictions are increasing with age.

Figure 5 – *Longitudinal and cross-sectional predicted probabilities of survival to age 75 and corresponding subjective probabilities by gender*



Male subjective survival probabilities to age 75 are broadly consistent with the longitudinal predictions, slightly lower until age 45 and slightly higher thereafter. Male subjective survival probabilities to age 90, instead, are always higher than the longitudinal predictions, and the difference between the two increases with age. On the other hand, female subjective survival probabilities do not exceed those of men and are always lower than the longitudinal predictions from the life tables. For survival to age 75, however, they are similar to the cross-sectional prediction based on the most recent (2003) life table.

Figure 6 – *Longitudinal and cross-sectional predicted probabilities of survival to age 90 and corresponding subjective probabilities by gender*



Figures 7 and 8 break down the analysis by geographic region (Northwest, Northeast, Center, and South/Islands). In the case of survival to age 75, the results are very similar to those obtained at the national level. In the case of survival to age 90, male subjective probabilities are similar to the longitudinal predictions for ages up to 40 years (except for the South), whereas female subjective probabilities are always lower than the longitudinal predictions, although the difference decreases with age.

4.2 Heterogeneity in survival probabilities

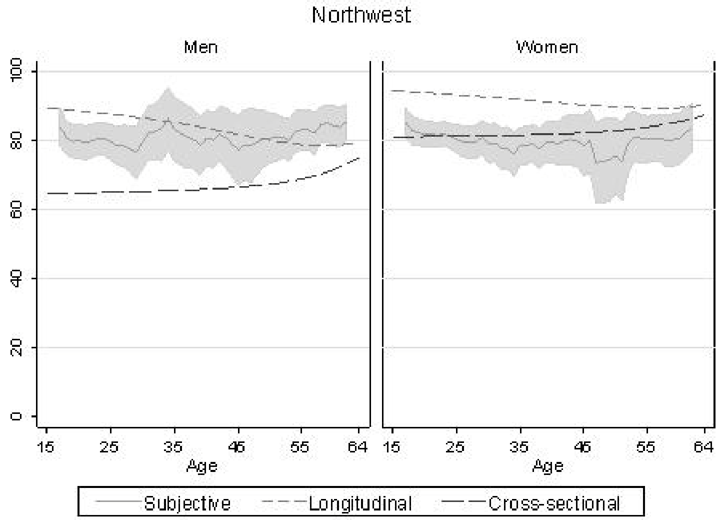
In the previous section we compared average subjective survival probabilities to predictions from cohort life tables. We now consider population heterogeneity in survival probabilities. Because the Italian life tables only provide information for coarsely defined population subgroups, this issue can only be studied using subjective data.

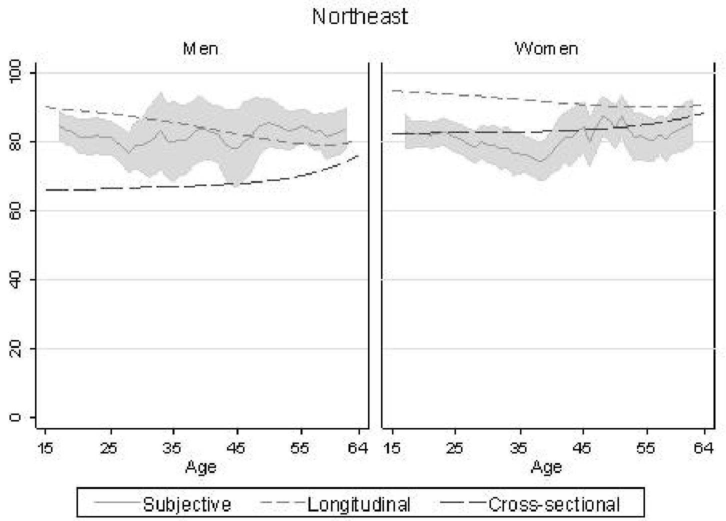
The natural approach for analyzing heterogeneity in subjective survival is regression. In this section we use regression methods to study how individual socio-demographic and economic characteristics help predict subjective probabilities of survival, focusing on survival to age 75. Our regression model may be regarded as a reduced-form production function for subjective longevity, of the type discussed by Benítez-Silva and Ni (2008). According to their approach, expected survival can be interpreted as the output of natural and biological initial conditions (proxied for example by age), decisions made up to the period in which the survival expectation is observed (proxied by education, marital status, activity status, type of employment, income, and reported health problems), and self-protection activities expected to have an effect on future survival (proxied by health care and other health investments).

Since subjective probabilities of survival to age 75 are reported on a 0-100 scale, and 36 percent of the sample answers either “0” or “100”, there is no simple transformation of reported probabilities that justifies the use of a linear model. In addition, the presence of both rounded values and focal responses suggests using a model for categorical data. Therefore, we estimate a number of ordered probit models by categorizing reported probabilities into the following eleven brackets: 0-5, 6-15, 16-25, 26-35, 36-45, 46-55, 56-65, 66-75, 76-85, 86-95, and 96-100.

Each model is estimated separately for men and women, first for the full sample and then for the subsample of people who are currently employed. The baseline respondent is an individual aged 40, with secondary education, currently employed, living in the Center, with no reported health problems, and married. For both men and women, the model is first estimated using only the basic information also available in the life tables, namely age and geographic region. We then include the binary indicators for education, in order to evaluate its effect in the most parsimonious model possible. Finally, we use the full set of available regressors (marital status, self-reported health status, and activity status). The same procedure is used for the subsample of employed people, for whom the full model also includes the logarithm of gross annual earnings and binary indicators for employment status, riskiness of the job, and satisfaction with the job characteristics[5].

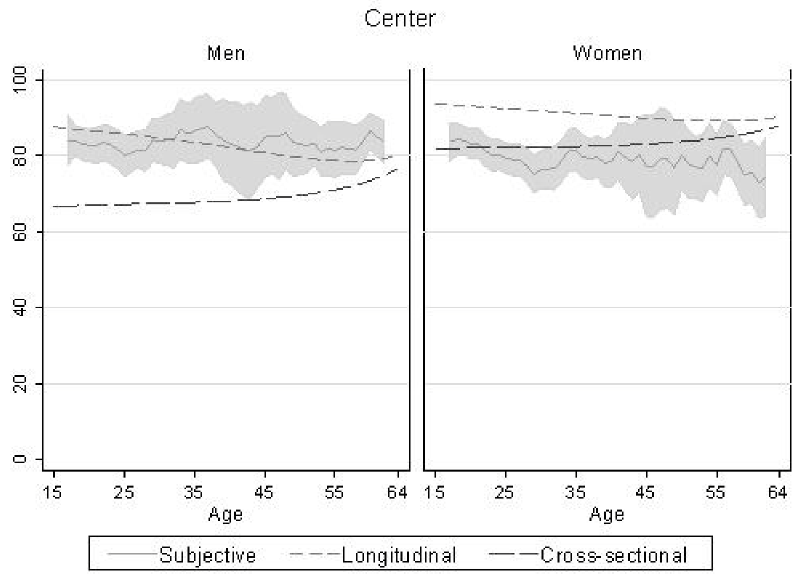
Figure 7 – *Longitudinal and cross-sectional predicted probability of survival to age 75 and corresponding subjective probabilities by gender and geographical region*





[5] (Log) earnings are rescaled by subtracting off their median value.

Figure 7 – cont'd



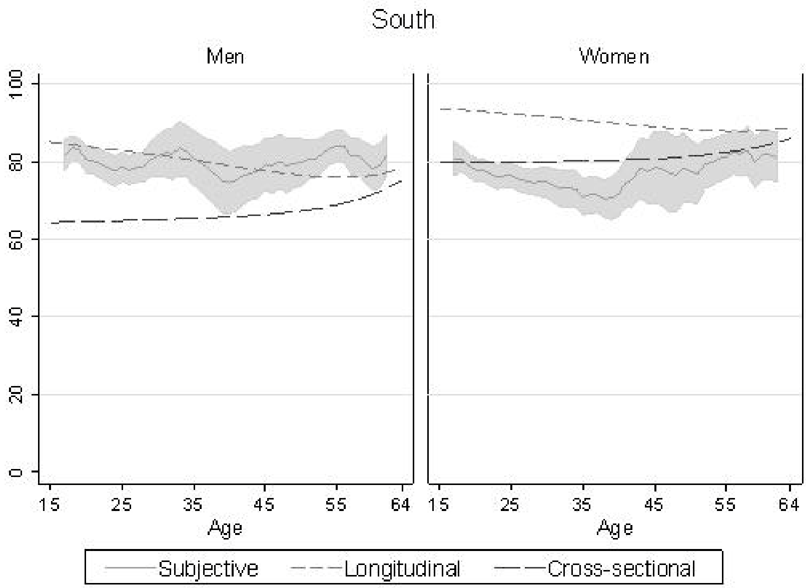
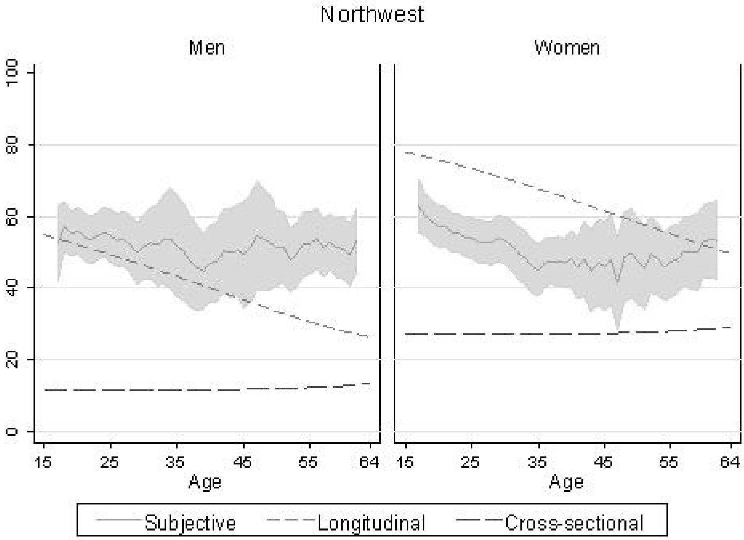


Figure 8 – *Longitudinal and cross-sectional predicted probability of survival to age 90 and corresponding subjective probabilities by gender and geographical region*



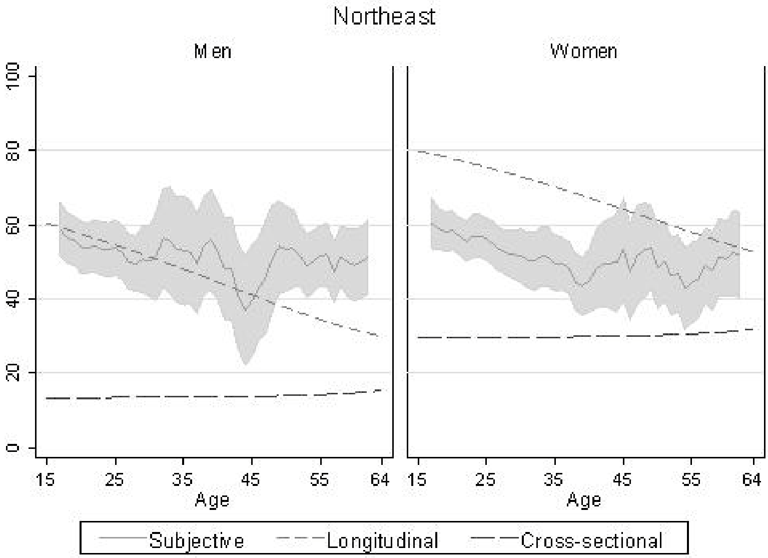
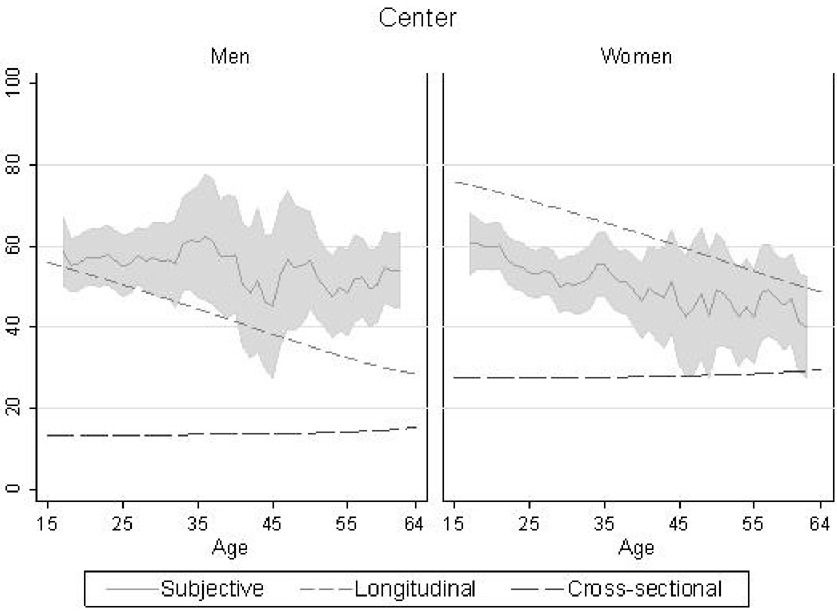
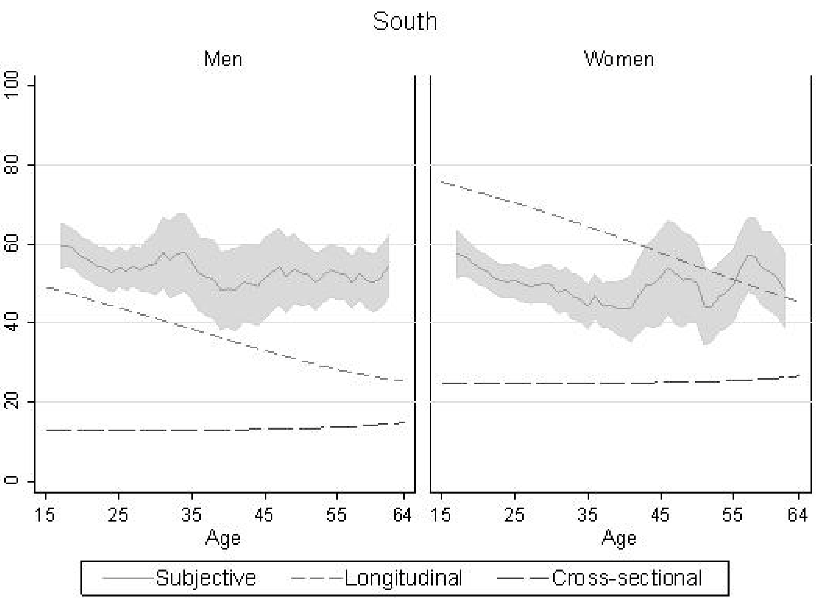


Figure 8 – cont'd





Results for men are presented in Table 4 and for women in Table 5 .The first three columns of each table refer to the full sample, whereas the last

Table 4 – *Ordered probit regressions for the probability of male survival to age 75*

Variable	All		
Age	0.002 ***	0.002 ***	0.001
Age squared	0.000 ***	0.000 ***	0.000 **
Region			
North	-0.058 **	-0.059 **	-0.063 **
South	-0.052 *	-0.053 *	-0.046
Education			
Primary or less	-	0.001	0.002
Tertiary	-	-0.022	-0.020
Not married	-	-	-0.077 **
Self-reported diseases			
Short term	-	-	-0.221 **
Long term	-	-	-0.429 ***
Activity status			
Unemployed	-	-	-0.089 ***
Inactive	-	-	0.042
Employment status			
Short term employee	-	-	-
Self-employed	-	-	-
Log annual earnings	-	-	-
Risky job	-	-	-
Satisfied with workload	-	-	-
Satisfied with safety	-	-	-
Satisfied with earnings	-	-	-
Satisfied with stability	-	-	-
No. obs.	10486	10486	10486
Log-likelihood	-18682.7	-18682.4	-18659.4
Pseudo R^2	0.001	0.001	0.002

Table 4 – cont'd

Variable	Employed			
Age	0.003 ***	0.004 ***	0.003 *	0.004 **
Age squared	0.000	0.000	0.000	0.000
Region				
North	-0.064	-0.063	-0.062	-0.067
South	-0.085 **	-0.084 *	-0.086 **	-0.061
Education				
Primary or less	-	-0.059	-0.060	-0.054
Tertiary	-	-0.042	-0.042	-0.050
Not married	-	-	-0.048	-0.049
Self-reported diseases				
Short term	-	-	-0.127	-0.073
Long term	-	-	-0.319 *	-0.282
Activity status				
Unemployed	-	-	-	-
Inactive	-	-	-	-
Employment status				
Short term employee	-	-	-	0.094 **
Self-employed	-	-	-	0.082 *
Log annual earnings	-	-	-	-0.013
Risky job	-	-	-	-0.107 ***
Satisfied with workload	-	-	-	0.069 *
Satisfied with safety	-	-	-	0.215 ***
Satisfied with earnings	-	-	-	0.084 ***
Satisfied with stability	-	-	-	0.127 ***
No. obs.	5139	5139	5139	5139
Log-likelihood	-9182.5	-9181.1	-9178.5	-9126.9
Pseudo R^2	0.001	0.001	0.001	0.007

Notes:* denotes asymptotic *p*-values between 5 and 10 percent, ** denotes asymptotic *p*-values between 1 and 5 percent, *** denotes asymptotic *p*-values below 1 percent.

Table 5 – *Ordered probit regressions for the probability of female survival to age 75*

Variable	All		
Age	0.002 ***	0.002 ***	0.003 ***
Age squared	0.001 ***	0.001 ***	0.001 ***
Region			
North	0.046 *	0.046 *	0.047 *
South	-0.122 ***	-0.122 ***	-0.127 ***
Education			
Primary or less	-	0.003	0.000
Tertiary	-	0.014	0.013
Not married	-	-	-0.024
Self-reported diseases			
Short term	-	-	-0.353 ***
Long term	-	-	-0.555 ***
Activity status			
Unemployed	-	-	0.054 *
Inactive	-	-	0.004
Employment status			
Short term employee	-	-	-
Self-employed	-	-	-
Log annual earnings	-	-	-
Risky job	-	-	-
Satisfied with workload	-	-	-
Satisfied with safety	-	-	-
Satisfied with earnings	-	-	-
Satisfied with stability	-	-	-
No. obs.	13171	13171	13171
Log-likelihood	-24212.1	-24211.9	-24190.4
Pseudo R^2	0.003	0.003	0.004

Table 5 – cont'd

Variable	Employed			
Age	0.004 ***	0.004 ***	0.003 **	0.004 ***
Age squared	0.000 ***	0.000 ***	0.001 ***	0.000 ***
Region				
North	0.074 *	0.074 *	0.075 *	0.069 *
South	-0.112 ***	-0.112 ***	-0.108 **	-0.088 **
Education				
Primary or less	-	0.004	0.004	0.012
Tertiary	-	0.003	0.008	0.029
Not married	-	-	-0.076 **	-0.064 *
Self-reported diseases				
Short term	-	-	-0.194	-0.157
Long term	-	-	-0.447 *	-0.397 *
Activity status				
Unemployed	-	-	-	-
Inactive	-	-	-	-
Employment status				
Short term employee	-	-	-	0.039
Self-employed	-	-	-	0.062
Log annual earnings	-	-	-	0.002
Risky job	-	-	-	-0.130 ***
Satisfied with workload	-	-	-	0.076 **
Satisfied with safety	-	-	-	0.110 ***
Satisfied with earnings	-	-	-	0.073 **
Satisfied with stability	-	-	-	0.119 ***
No. obs.	5038	5038	5038	5038
Log-likelihood	-9385.3	-9385.3	-9380.6	-9347.3
Pseudo R^2	0.002	0.002	0.003	0.006

Notes:* denotes asymptotic *p*-values between 5 and 10 percent, ** denotes asymptotic *p*-values between 1 and 5 percent, *** denotes asymptotic *p*-values below 1 percent.

four columns refer to the subsample of people who are currently employed. Most results are similar for men and women: in all models, the age coefficient is positive, and larger in models containing more covariates, as found respectively by Hurd and McGarry (1995) and Elder (2007). Subjective survival probabilities are lower for men or women who live in the South, are not married, or report having health problems[6]. Somewhat surprisingly, educational attainments do not appear to matter, possibly because of attenuation bias due to measurement error in self-reported

[6] The higher mortality among widows has been documented by demographers since Farr (1858). Recently, Elwert and Christakis (2008) have provided evidence in support of a causal interpretation of the so-called "widowhood effect".

education. As suggested by one referee, this result may also reflect a feature of the Italian case, where returns to education are low and health care is universal. As a consequence, education and income may not affect health outcomes, all else equal. Among the currently employed, subjective survival probabilities are lower for those who hold a risky job, and higher for those who say they are satisfied with job characteristics such as compliance with safety regulations, earnings, and job stability. Annual earnings do not seem to matter. Subjective survival probabilities are also lower among men living in the North or unemployed, and higher for males who are self-employed or have a short-term employment contract.

4.3 Controlling for item nonresponse

Our results in Section 4.2 are somewhat surprising because they show no evidence of a relationship between education and subjective survival probabilities, even before controlling for health status. This contrasts with previous evidence of a negative effect of education on both observed mortality (Lleras-Muney, 2005) and subjective probabilities of survival (Hurd and McGarry, 1995; Benítez-Silva and Ni, 2008). An important feature of our data that could affect our results is the large fraction of individuals (about one third) who did not answer the questions on subjective survival probabilities. To check whether this is the case, we also estimate an ordered probit model that takes into account sample selection due to item nonresponse[7].

Table 6 presents the results of this model, separately for men and women. The first column reports for comparison the estimates of the naïve ordered probit model for the subsample of nonmissing observations[8], the second column reports the estimated coefficients for the selection equation, whereas the third column presents the estimates for the ordered probit that takes sample selection into account. All observations are included in the analysis (16,825 men and 20,688 women)[9]. The covariates in the selection equation are all the variables which are available for all observations, plus three variables that we assume affect nonresponse but not subjective survival probabilities, namely binary indicators for being able to speak English, being able to use a personal computer, and being a panel respondent. All these indicators have a strongly significant positive relationship with the probability

[7] The log-likelihood of the model is reported in Appendix B.

[8] The results for the naïve model are reported in the third column of Table 4 for men and of Table 5 for women.

[9] Results do not change if the model with sample selection is estimated using the subset of individuals for whom all covariates are available but subjective survival probabilities may be missing.

Table 6 – *Ordered probit regression with sample selection for the probability of survival to age 75*

Variable	Naïve model	Sample selection model	
	S_{75}	Selection	S_{75}
		Men	
Age	0.001	-0.015 ***	0.000
Age squared	0.000 **	0.000 *	0.000 **
Region			
North	-0.063 **	0.072 **	-0.061 **
South	-0.046	-0.049 *	-0.048
Education			
Primary or less	0.002	-0.180 ***	-0.003
Tertiary	-0.020	-0.014	-0.019
Not married	-0.077 **	-0.098 ***	-0.079 **
Self-reported diseases			
Short term	-0.221 **	-0.308 ***	-0.230 **
Long term	-0.429 ***	-0.395 ***	-0.440 ***
Activity status			
Unemployed	-0.089 ***	0.037	-0.090 ***
Inactive	0.042	0.207 ***	0.047
English speaker	-	0.037	-
Panel respondent	-	0.229 ***	-
PC user	-	0.180 ***	-
Constant	-	0.114 ***	-
ρ	-	-	0.043
No. obs.	10486	16825	
Log-likelihood	-18659.4	-29288.2	

Table 6 – cont'd

Variable	Naïve model S_{75}	Sample selection model Selection	Sample selection model S_{75}
		Women	
Age	0.003 ***	-0.015 ***	0.001
Age squared	0.001 ***	0.000	0.001 ***
Region			
North	0.047 *	0.039	0.049 *
South	-0.127 ***	-0.095 ***	-0.135 ***
Education			
Primary or less	0.000	-0.150 ***	-0.015
Tertiary	0.013	0.034	0.018
Not married	-0.024	-0.021	-0.026
Self-reported diseases			
Short term	-0.353 ***	-0.272 ***	-0.377 ***
Long term	-0.555 ***	-0.393 ***	-0.587 ***
Activity status			
Unemployed	0.054 *	0.097 ***	0.055 **
Inactive	0.004	0.120 ***	0.011
English speaker	-	0.053 **	-
Panel respondent	-	0.245 ***	-
PC user	-	0.106 ***	-
Constant	-	0.033	-
P	-	-	0.150 *
No. obs.	13171	20688	
Log-likelihood	-24190.4	-37230.6	

Notes:* denotes asymptotic *p*-values between 5 and 10 percent, ** denotes asymptotic *p*-values between 1 and 5 percent, *** denotes asymptotic *p*-values below 1 percent.

of being included in the sample. Individuals are also more likely to be included if they are younger, more educated, without health problems, and unemployed or inactive. However, the correlation coefficient ρ is not statistically different from zero, and the coefficients of the ordered probit model do not change much compared to the coefficient in the naïve model, except for health-related variables for women. Again, educational attainments do not seem to matter.

5. SUMMARY AND CONCLUSIONS

Subjective survival probabilities are increasingly used by researchers to understand mortality differentials due to biological conditions, socio-economic status, and health-related behavior. Because life tables only provide mortality rates for broadly defined population subgroups, they cannot be used for such purposes. An important preliminary issue, however, is whether subjective survival probabilities are on average consistent with life tables. Previous comparisons provide limited and inconclusive evidence, since they use life tables without properly taking into account the presence of strong cohorts effects. This paper provides more convincing evidence by comparing average subjective survival probabilities from a representative sample of individuals aged 15-64 with the survival probabilities obtained from cohort life tables constructed using a long sequence of cross-sectional life tables.

We find that male subjective survival probabilities are reasonably close to the longitudinal predictions from life tables, particularly for survival to age 75, whereas female subjective survival probabilities are always lower than the life-table predictions. This result is consistent with previous evidence on gender differentials in self-assessed health, which suggests that women tend to underestimate their actual survival probabilities. An alternative view, also consistent with our results, is that own survival is best predicted by individuals themselves (Perozek, 2008), and therefore life table predictions of female life expectancy should be revised downward.

Expected survival can be viewed as the output of a production function with biological conditions, health shocks, and self-protection activities as inputs. While life tables only provide information on a limited set of these dimensions (age, cohort, gender, and geographical area), survey data offer the possibility of studying the relationship between subjective survival probabilities and a large set of individual characteristics, including self-reported health, marital status, activity status, and job characteristics. We improve on the current literature by using a discrete choice model (which partially addresses the issue of focal responses), and by controlling for potential sample selection bias due to item nonresponse. Most of our results confirm previous findings except that, somewhat surprisingly, we find no statistically significant relationship between educational attainments and subjective survival probabilities.

Acknowledgements

We thank Graziella Caselli and Emiliano Mandrone for insightful discussions, and an anonymous referee for helpful comments.

References

BALIA S. (2007), "Reporting expected longevity and smoking: Evidence from the SHARE", *HEDG Working Paper*, N°. 07/10, University of York.

BENITEZ-SILVA H., NI H. (2008), "Health status and health dynamics in an empirical model of expected longevity", *Journal of Health Economics*, 27, 564-584.

BRUINE DE BRUIN W., FISCHBECK P.S., STIBER N.A., FISCHHOFF B. (2002), "What number is 'fifty-fifty'?: Redistributing excessive 50% responses in elicited probabilities", *Risk Analysis*, 22, 713-723.

CASE A., PAXSON C. (2005), "Sex differences in morbidity and mortality", *Demography*, 42, 189-214.

DELAVANDE A., ROHWEDDER S. (2008), "Differential mortality in Europe and the U.S.: Estimates based on subjective probabilities of survival", *RAND Working Paper*, No. 613.

EHRLICH I. (2000), "Uncertain lifetime, life protection, and the value of life saving", *Journal of Health Economics*, 19, 341-367.

ELDER T.E. (2007), "Subjective survival probabilities in the Health and Retirement Study: Systematic biases and predictive validity", Michigan Retirement Research Center, *Working Paper*, No. 159.

ELWERT F., CHRISTAKIS N.A. (2008), "Wives and ex-wives: A new test for homogamy bias in the widowhood effect", *Demography*, 45, 851-873.

FARR W. (1858), "Influence of marriage on the mortality of the French people", in HASTINGS G.W. (ed.), *Transactions of the National Association for the Promotion of Social Science*, Park, London.

GROSSMAN M. (1972), "On the concept of health capital and the demand for health", *Journal of Political Economy*, 80, 223-255.

GUISO L., TISENO A., WINTER J. (2005), "Expectations", in BÖRSCH-SUPAN, A., BRUGIAVINI A., JÜRGES H., MACKENBACH J., SIEGRIST J., WEBER G. (eds.), *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*, MEA, Mannheim.

HAMERMESH D.S. (1985), "Expectations, life expectancy, and economic behavior", *Quarterly Journal of Economics*, 100, 389-408.

HILL, D., PERRY, M., WILLIS, R.J. (2004) "*Estimating Knightian Uncertainty from Survival Probability Questions on the HRS*", paper prepared for the conference "*Labor and Social Interaction in Economics*", December 16-17, 2004, The Eitan Berglas School of Economics, Tel-Aviv University. http://www.aeaweb.org/annual_mtg_papers/2005/0109_13001003.pdf.

HURD M., McGARRY K. (1995), "Evaluation of the subjective probabilities of survival in the Health and Retirement Study", *The Journal of Human Resources*, 30, S268-S292.

HURD M., MCGARRY K. (2002), "The predictive validity of subjective probabilities of survival", *Economic Journal*, 112, 966-985.

ISFOL (2006), *PLUS Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro*, Rubbettino, Soveria Mannelli (Catanzaro).

LEE R.D., CARTER L.R. (1992), "Modeling and forecasting U.S. mortality", *Journal of the American Statistical Association*, 87, 659-671.

LILLARD, L.A., WILLIS R.J. (2001), "*Cognition and wealth: The importance of probabilistic thinking*", Michigan Retirement Research Center Working Paper No. 007. http://www.mrrc.isr.umich.edu/publications/papers/pdf/wp007.pdf.

LLERAS-MUNEY A. (2005), "The relationship between education and adult mortality in the United States", *Review of Economic Studies*, 72, 189-221.

NATHANSON C. (1975), "Illness and the feminine role: A theoretical review", *Social Science and Medicine*, 9, 57-62.

PEROZEK M. (2008), "Using subjective expectations to forecast longevity: Do survey respondents know something we don't know?", *Demography*, 45, 95-113.

SMITH V.K., TAYLOR D.H., SLOAN F.A. (2001), "Longevity expectations and death: Can people predict their own demise?", *American Economic Review*, 91, 1126-1134.

WANG D. (2000), "Modeling mortality data using the Lee-Carter model", *Stata Technical Bulletin*, 57, 15-17.

Appendix A
The PLUS survey

The Participation, Labour and Unemployment Survey (PLUS) is a cross-sectional survey with a longitudinal component carried out by ISFOL (Istituto per lo lo Sviluppo della Formazione Professionale dei Lavoratori). The reference population consists of individuals aged 15-64 living in a private household, excluding some portions identified by age, gender, and activity status. The only relevant excluded group consists of women aged 40-64, who are inactive but not retired (11.3 percent of the population aged 15-64). The other groups account for less than 3 percent of the population. The sample was selected by quota sampling, taking into account region and size of municipalities, age, gender, and labor force status (ISFOL, 2006). Interviews were carried out by computer-assisted telephone interview, excluding proxy respondents.

The PLUS questionnaire is organized in modules. Some of them (demographics, labor force status, education and training, *etc.*) are asked to everybody, whereas others are devoted only to particular target groups (young people aged 15-29 years, women aged 20-49 years, people aged 50-64 years, people looking for a job, and people with a job). The dataset includes detailed information on education, training, and job characteristics. Unfortunately, very little is known about household members other than the interviewee: only their gender, birth year, activity status, and relationship to the household head (which is not necessarily the respondent). As for the respondent's parents, their educational attainments and main occupation is available. Household income or wealth is not available, nor is information on individual unearned income.

As for other individual characteristics, the survey collects information on age, gender, geographical region and town size, educational attainments, marital status, presence of children, household head status, and nationality. Self-reported health status is classified into 3 categories: no health problems, short term health problems (lasting less than 6 months or occasional), and long term health problems (lasting or expected to last more than 6 months). As for activity status, it is possible to distinguish between employment, unemployment, retirement, and other types of inactivity. For individuals who are currently working, several job characteristics are available, such as the type of employment (long term dependent employment, short term dependent employment, and self-employment), gross annual earnings, usual weekly hours, satisfaction with specific aspects of the job (workload, compliance with safety regulations, earnings, and stability), and whether the worker thinks her job is risky for her own health. As for subjective survival probabilities, the wording of the question in the PLUS questionnaire is:

For scientific purposes only, we would like to ask you:
"In your opinion, what is the probability that you will reach age 75, and age 90?" Please provide a value between 100 (certain event) and 0 (impossible event).

Appendix B
Log-likelihood of an ordered probit model with sample selection

Let Y^* denote a latent random variable, defined on the whole real line, which depends linearly on a set of covariates X and a random error ε, that is, $Y^* = X\beta + \varepsilon$. Instead of Y^*, we observe an ordered categorical random variable Y which is equal to j whenever $k_{j-1} < Y^* \leq k_j$, with $k_0 = -\infty$ and $k_J = \infty$. In turn, Y is only observed whenever $Z^* > 0$, where Z^* is another latent random variable which depends linearly on a set of covariates W and a random error u, that is, $Z^* = W\gamma + u$. Finally, let Z be a binary indicator equal to 1 if $Z^* > 0$ and equal to 0 otherwise. In our paper, $J = 11$ and we have the following table:

S_{75}	Y	k_j	k_{j+1}	Z
0 – 5	1	$-\infty$	k_1	1
6 – 15	2	k_1	k_2	1
16 – 25	3	k_2	k_3	1
26 – 35	4	k_3	k_4	1
36 – 45	5	k_4	k_5	1
46 – 55	6	k_5	k_6	1
56 – 65	7	k_6	k_7	1
66 – 75	8	k_7	k_8	1
76 – 85	9	k_8	k_9	1
86 – 95	10	k_9	k_{10}	1
96 – 100	11	k_{10}	$+\infty$	1
–	–	–	–	0

Let the bivariate distribution of the random errors ε and u be normal (Gaussian) with zero means, unit variances and correlation coefficient ρ, and let $\phi_2(\cdot,\cdot,\rho)$ and $\Phi_2(\cdot,\cdot,\rho)$ respectively denote the density and the distribution function of such distribution. Denote by $\Phi(\cdot)$ the distribution function of a univariate standard normal and by $I\{A\}$ the indicator function of the event A. Then the contribution to the log-likelihood of a single observation i is:

$$\begin{aligned}\log L_i \quad &= Z_i \sum_{j=1}^{J} I\{Y_i = j\} \log\left[\int_{k_j - \beta' X_i}^{k_{j+1} - \beta' X_i} \int_{-\gamma' W_i}^{+\infty} \phi_2(\varepsilon, u, \rho)\, du\, d\varepsilon\right] + (1 - Z_i) \log[1 - \Phi(\gamma' W_i)] \\ &= Z_i \sum_{j=1}^{J} I\{Y_i = j\} \log\left[\int_{-\infty}^{k_{j+1} - \beta' X_i} \int_{-\infty}^{\gamma' W_i} \phi_2(\varepsilon, u, -\rho)\, du\, d\varepsilon\right. \\ &\qquad \left. - \int_{-\infty}^{k_j - \beta' X_i} \int_{-\infty}^{\gamma' W_i} \phi_2(\varepsilon, u, -\rho)\, du\, d\varepsilon\right] + (1 - Z_i) \log[1 - \Phi(\gamma' W_i)] \\ &= Z_i \sum_{j=1}^{J} I\{Y_i = j\} \log\left[\Phi_2\left(k_{j+1} - \beta' X_i, \gamma' W_i, -\rho\right) - \Phi_2\left(k_j - \beta' X_i, \gamma' W_i, -\rho\right)\right] \\ &\qquad + (1 - Z_i) \log[1 - \Phi(\gamma' W_i)]\end{aligned}$$