Appendix

Probability Mass Functions

Probability mass functions are defined recursively. This appendix contains recursions for *YI*, *YFS*, *YIR*, and *YAL*. It also contains definitions of conditional years in retirement (*CYIR*) and conditional years to retirement (*CYTR*) which are conditional of a person spending a positive amount of time in retirement. The recursions for *YA* are already in the literature (Skoog and Ciecka, 2012). Recursions transform probabilities at age x+1into probabilities at age x. They are backwards in the sense that they begin at age TA–1and move to successively younger ages.

Recursions defining probability mass functions for random variables consist of global conditions, boundary conditions, and main recursions. Global conditions apply to all random variables. Boundary conditions deal with probabilities that random variables take on values 0 or .5 years (*i.e.*, the smallest values for random variables *YA*, *YI*, *YFS*, and *YIR*). Main recursions capture probabilities for all other values of *YA*, *YI*, *YFS*, and *YIR*.

Global conditions are specified in (1.1)-(1.5). They consist of the following statements: (1.1) no random variable can be negative or exceed $TA-x-.5$ years for a person age x; (1.2) the value of all random variables must be zero at age TA; (1.3) the probability of dying does not depend of activity/inactivity states at ages and x–1 and x; (1.4) at age TA – 1, the only transition is to death which occurs at age TA – .5 or, said differently, transitions to either a or i occur with probability zero; and (1.5) given labor force statuses at ages x–1 and x, transitions to activity, inactivity, and death must sum to 1.00.

Global Conditions for Probability Mass Functions

$$p_{RV}(x,m,n,y)=0 \text{ for } y<0 \text{ or } y>TA-x-.5 \text{ for } RV = YA, YI, YFS, YIR \tag{1.1}$$

$$p_{RV}(TA,m,n,0)=1 \text{ for } RV = YA, YI, YFS, YIR \tag{1.2}$$

$${}^{mn}p_x^d = {}^{\bullet}p_x^d \quad \text{for } x = BA,\ldots,TA-1 \tag{1.3}$$

$${}^{mn}p_{TA-1}^d = 1 \text{ and } {}^{mn}p_{TA-1}^m = {}^{mn}p_{TA-1}^n = 0 \tag{1.4}$$

$${}^{aa}p_x^a + {}^{aa}p_x^i + {}^{\bullet}p_x^d = 1; \quad {}^{ai}p_x^a + {}^{ai}p_x^i + {}^{\bullet}p_x^d = 1 \tag{1.5}$$

$${}^{ia}p_x^a + {}^{ia}p_x^i + {}^{\bullet}p_x^d = 1; \quad {}^{ii}p_x^a + {}^{ii}p_x^i + {}^{\bullet}p_x^d = 1 \text{ for } x = BA,\ldots,TA-1$$

Recursions for *YA*, *YI*, *YFS*, and *YIR* transform probabilities at age x+1into probabilities at age x. We explain the underlying intuition of the recursions using *YI* as an example. Equations (2.1)-(2.5) are boundary conditions for the probability of 0 or .5 years in inactivity. In (2.1), when inactive at age x, there is a zero probability of no future inactivity because the earliest transition out of inactivity occurs at age x+.5. Therefore, at least .5 years of future inactivity accrue to a person inactive at age x, regardless of activity status at age x–1, as shown in equations (2.4) and (2.5). This occurs through a transition to activity and staying active thereafter or dying at age x+.5. In a similar manner, people active at age x accrue no future inactive time by always staying active in the future or by dying while active at age x+.5 as shown in (2.2) and (2.3). Boundary conditions (3.1)-(3.8) for *YFS* and (4.1)-(4.8) for *YIR* work in a similar manner as those for *YI*.

The fundamental idea in the main recursions relates to the manner in which values of *YA*, *YI*, *YFS*, and *YIR* at age x+1 relate to values at age x. For example, the first term on the right hand side of *YI* recursion (2.9) is ${}^{ai}p_x^i\, p_{YI}\left(x+1,i,i,y-1\right)$. The first factor in this term is the transition probability ${}^{ai}p_x^i$: the probability that a person active at age x–1, inactive at age x, remains inactive at age x +1. The second factor $p_{YI}\left(x+1,i,i,y-1\right)$ denotes the probability that a person age x+1 who was inactive at age x and inactive at age x+1 accumulates y–1 years of future inactivity. The product ${}^{ai}p_x^i\, p_{YI}\left(x+1,i,i,y-1\right)$ contributes to the probability $p_{YI}\left(x,a,i,y\right)$ that a person age x who was active at age x–1 and inactive at age x will accumulate $y=(y-1)+1$ years of inactivity – *i.e.*, the probability of y–1 years of inactivity at age x+1 becomes the probability of y years of inactivity at age x; but this occurs only with probability ${}^{ai}p_x^i$ of remaining inactive between ages x and x+1, hence the first term ${}^{ai}p_x^i\, p_{YI}\left(x+1,i,i,y-1\right)$ on the right side of (2.9). The second term on the right hand side of recursion (2.9) is ${}^{ai}p_x^a\, p_{YI}\left(x+1,i,a,y-.5\right)$. The first factor in this term is the transition probability ${}^{ai}p_x^a$: the probability that a person active at age x–1, inactive at age x, is active at age x+1, with the transition to activity occurring at age x+.5. This means there is .5 of a year of inactivity between ages x and x+.5. The second factor $p_{YI}\left(x+1,i,a,y-.5\right)$ denotes the probability that a person age x+1 who was inactive at age x and active at age x+1 accumulates y–.5 years of future inactivity. The probability of y–.5 future years of inactivity at age x+1 becomes the probability of $y=(y-.5)+.5$ years of inactivity at age x; but this occurs only with probability ${}^{ai}p_x^a$. Thus, the product ${}^{ai}p_x^a\, p_{YI}\left(x+1,i,a,y-.5\right)$ contributes to the probability $p_{YI}\left(x,a,i,y\right)$. Finally, adding ${}^{ai}p_x^i\, p_{YI}\left(x+1,i,i,y-1\right)$ and ${}^{ai}p_x^a\, p_{YI}\left(x+1,i,a,y-.5\right)$, we get *YI* recursion (2.9) for $p_{YI}\left(x,a,i,y\right)$. Other main recursions for *YI* and main recursions for *YFS* and *YIR* work in a similar manner as recursion (2.9) by transforming probabilities at age x+1 to probabilities at age x using transition probabilities as weights.

Boundary Conditions for Probability Mass Functions for $YI_{x,m,n}=y$

$$p_{YI}\left(x,i,i,0\right)=p_{YI}\left(x,a,i,0\right)=0 \tag{2.1}$$

$$p_{YI}\left(x,i,a,0\right)={}^{ia}p_x^a\, p_{YI}\left(x+1,a,a,0\right)+{}^{ia}p_x^d \tag{2.2}$$

$$p_{YI}\left(x,a,a,0\right)={}^{aa}p_x^a\, p_{YI}\left(x+1,a,a,0\right)+{}^{aa}p_x^d \tag{2.3}$$

$$p_{YI}\left(x,i,i,.5\right)={}^{ii}p_x^a\,p_{YI}\left(x+1,i,a,0\right)+{}^{ii}p_x^d \tag{2.4}$$
$$p_{YI}\left(x,a,i,.5\right)={}^{ai}p_x^a\,p_{YI}\left(x+1,i,a,0\right)+{}^{ai}p_x^d \tag{2.5}$$
for $x = BA,\ldots,TA-1$

Main Recursions for Probability Mass Functions for $YI_{x,m,n} = y$

$$p_{YI}\left(x,i,a,y\right)={}^{ia}p_x^i\,p_{YI}\left(x+1,a,i,y-.5\right)+{}^{ia}p_x^a\,p_{YI}\left(x+1,a,a,y\right) \tag{2.6}$$
$$p_{YI}\left(x,a,a,y\right)={}^{aa}p_x^i\,p_{YI}\left(x+1,a,i,y-.5\right)+{}^{aa}p_x^a\,p_{YI}\left(x+1,a,a,y\right) \tag{2.7}$$
for $y = 1,2,3,\ldots,TA-x-1$ and $x = BA,\ldots,TA-2$
$$p_{YI}\left(x,i,i,y\right)={}^{ii}p_x^i\,p_{YI}\left(x+1,i,i,y-1\right)+{}^{ii}p_x^a\,p_{YI}\left(x+1,i,a,y-.5\right) \tag{2.8}$$
$$p_{YI}\left(x,a,i,y\right)={}^{ai}p_x^i\,p_{YI}\left(x+1,i,i,y-1\right)+{}^{ai}p_x^a\,p_{YI}\left(x+1,i,a,y-.5\right) \tag{2.9}$$
for $y = 1.5,2.5,3.5,\ldots,TA-x-.5$ and $x = BA,\ldots,TA-2$

Boundary Conditions for Probability Mass Functions for $YFS_{x,m,n} = y$

$$p_{YFS}(x,a,a,y)=0 \tag{3.1}$$
$$p_{YFS}(x,i,a,y)=0 \tag{3.2}$$
for $y = 0,1,2,3,\ldots,TA-1$ and $x = BA,\ldots,TA-1$
$$p_{YFS}(x,a,i,y)=0 \tag{3.3}$$
$$p_{YFS}(x,i,i,y)=0 \tag{3.4}$$
for $y = .5,1,2,3,\ldots,TA-1$ and $x = BA,\ldots,TA-1$
$$p_{YFS}(x,a,a,.5)={}^{aa}p_x^d+{}^{aa}p_x^i\,p_{YFS}(x+1,a,i,0) \tag{3.5}$$
$$p_{YFS}(x,i,a,.5)={}^{ia}p_x^d+{}^{ia}p_x^i\,p_{YFS}(x+1,a,i,0) \tag{3.6}$$
$$p_{YFS}(x,a,i,0)={}^{ai}p_x^d+{}^{ai}p_x^i\,p_{YFS}(x+1,i,i,0) \tag{3.7}$$
$$p_{YFS}(x,i,i,0)={}^{ii}p_x^d+{}^{ii}p_x^i\,p_{YFS}(x+1,i,i,0) \tag{3.8}$$
for $x = BA,\ldots,TA-1$

Main Recursions for Probability Mass Functions for $YFS_{x,m,n} = y$

$$p_{YFS}(x,a,a,y)={}^{aa}p_x^a\,p_{YFS}(x+1,a,a,y-1)+{}^{aa}p_x^i\,p_{YFS}(x+1,a,i,y-1) \tag{3.9}$$
$$p_{YFS}(x,i,a,y)={}^{ia}p_x^a\,p_{YFS}(x+1,a,a,y-1)+{}^{ia}p_x^i\,p_{YFS}(x+1,a,i,y-1) \tag{3.10}$$
$$p_{YFS}(x,a,i,y)={}^{ai}p_x^a\,p_{YFS}(x+1,i,a,y-1)+{}^{ai}p_x^i\,p_{YFS}(x+1,i,i,y-1) \tag{3.11}$$
$$p_{YFS}(x,i,i,y)={}^{ii}p_x^a\,p_{YFS}(x+1,i,a,y-1)+{}^{ii}p_x^i\,p_{YFS}(x+1,i,i,y-1) \tag{3.12}$$
for $y = 1.5, 2.5, 3.5,\ldots,TA-x-.5$ and for $x = BA,\ldots,TA-2$

Boundary Conditions for Probability Mass Functions for $YIR_{x,m,n} = y$

$$p_{YIR}(x,a,a,0)={}^{aa}p_x^a\,p_{YIR}(x+1,a,a,0)+{}^{aa}p_x^i\,p_{YIR}(x+1,a,i,0)+{}^{aa}p_x^d \tag{4.1}$$
$$p_{YIR}(x,i,a,0)={}^{ia}p_x^a\,p_{YIR}(x+1,a,a,0)+{}^{ia}p_x^i\,p_{YIR}(x+1,a,i,0)+{}^{ia}p_x^d \tag{4.2}$$
$$p_{YIR}(x,a,i,0)={}^{ai}p_x^a\,p_{YIR}(x+1,i,a,0)+{}^{ai}p_x^i\,p_{YIR}(x+1,i,i,0) \tag{4.3}$$

$$p_{YIR}(x,i,i,0) = {}^{ii}p_x^a\, p_{YIR}(x+1,i,a,0) + {}^{ii}p_x^i\, p_{YIR}(x+1,i,i,0) \tag{4.4}$$

$$p_{YIR}(x,a,a,y) = 0\,,\quad y = .5,1.5,2.5,3.5,\ldots,TA\text{–}x\text{–}.5 \tag{4.5}$$

$$p_{YIR}(x,i,a,y) = 0\,,\quad y = .5,1.5,2.5,3.5,\ldots,TA\text{–}x\text{–}.5 \tag{4.6}$$

$$p_{YIR}(x,a,i,.5) = {}^{ai}p_x^d \tag{4.7}$$

$$p_{YIR}(x,i,i,.5) = {}^{ii}p_x^d \tag{4.8}$$

for $x = BA,\ldots.,TA - 1$

Main Recursions for Probability Mass Functions for $YIR_{x,m,n} = y$

$$p_{YIR}(x,a,i,y) = {}^{ai}p_x^i\, p_{YIR}(x+1,i,i,y-1) \tag{4.9}$$

$$p_{YIR}(x,i,i,y) = {}^{ii}p_x^i\, p_{YIR}(x+1,i,i,y-1) \tag{4.10}$$

for $y = 1.5,2.5,\ldots,TA\text{–}x\text{–}.5$ and $x = BA,\ldots,TA\text{–}2$

$$p_{YIR}(x,a,a,y) = {}^{aa}p_x^a\, p_{YIR}(x+1,a,a,y) + {}^{aa}p_x^i\, p_{YIR}(x+1,a,i,y-.5) + {}^{aa}p_x^i\, p_{YIR}(x+1,a,i,y) \tag{4.11}$$

$$p_{YIR}(x,i,a,y) = {}^{ia}p_x^a\, p_{YIR}(x+1,a,a,y) + {}^{ia}p_x^i\, p_{YIR}(x+1,a,i,y-.5) + {}^{ia}p_x^i\, p_{YIR}(x+1,a,i,y) \tag{4.12}$$

$$p_{YIR}(x,a,i,y) = {}^{ai}p_x^a\, p_{YIR}(x+1,i,a,y) + {}^{ai}p_x^i\, p_{YIR}(x+1,i,i,y) \tag{4.13}$$

$$p_{YIR}(x,i,i,y) = {}^{ii}p_x^a\, p_{YIR}(x+1,i,a,y) + {}^{ii}p_x^i\, p_{YIR}(x+1,i,i,y) \tag{4.14}$$

for $y = 1,2,3,\ldots,TA\text{–}x\text{–}1$ and $x = BA,\ldots,TA\text{–}2$

Probability mass functions for *YAL* are defined by the following boundary conditions and main recursions.

Boundary Conditions

$$p_{YAL}(x,a,a,.5) = p_{YAL}(x,i,a,.5) = p_{YAL}(x,a,i,.5) = p_{YAL}(x,i,i,.5) = p_{YAL}(x,.5) = {}^{\bullet}p_x^d \tag{5.1}$$

for $x = BA,\ldots,TA - 1$

Main Recursions

$$p_{YAL}(x,a,a,y) = p_{YAL}(x,i,a,y) = p_{YAL}(x,a,i,y) = p_{YAL}(x,i,i,y)$$
$$= p_{YAL}(x,y) = (1 - {}^{\bullet}p_x^d)\, p_{YAL}(x+1,y-1) \tag{5.2}$$

for $y = 1.5,2.5,3.5,\ldots,TA - x - .5$; and $x = BA,\ldots,TA - 1$.

Complete life expectancy can be decomposed using $YAL_x = YA_{x,m.n} + YI_{x,m,n}$. For every realization in the underlying probability space, the sum of the two random variables is years lived. Taking expectations, the sum of expected years of activity and expected years of inactivity equals complete life expectancy $\dot{e}_x \equiv E[YAL_x] = E[YA_{x,m.n}] + E[YI_{x,m,n}]$. It also is the case that $YAL_x \equiv YFS_{x,m.n} + YIR_{x,m,n}$ and $\dot{e}_x \equiv E[YAL_x] = E[YFS_{x,m.n}] + E[YIR_{x,m,n}]$.

We observe that $E[YIR_{x,m,n}] \le E[YI_{x,m,n}]$ because all time in retirement is inactive time, but not all inactivity occurs in retirement. The ratio $0 \le E[YIR_{x,m,n}] / E[YI_{x,m,n}] \le 1$ tells us the fraction of future inactive time spent in retirement. This ratio is low if people tend to leave the labor force for substantial periods of time, reenter the labor force for a brief period of time, and then die while active or soon after once again turning inactive.

Boundary conditions and main recursions (4.1)-(4.14) allow for the possibility that a person may die while active and therefore spend no time in retirement. In such a case, $YIR_{x,m,n} = 0$. While this probability is of interest in itself because it defines the probability of retirement never occurring, we define the conditional years-in-retirement random variable *CYIR* as years in retirement conditioned on $YIR > 0$ with pmfs defined in (6.1)-(6.4)

$$p_{CYIR}(x,a,i,y) = \frac{p_{YIR}(x,a,i,y)}{\sum_{y=.5.1.0,1.5...}^{TA-x-.5} p_{YIR}(x,a,i,y)} \tag{6.1}$$

$$p_{CYIR}(x,i,i,y) = \frac{p_{YIR}(x,i,i,y)}{\sum_{y=.5,1.0.1.5,...}^{TA-x-.5} p_{YIR}(x,i,i,y)} \tag{6.2}$$

for $y = .5, 1.0, 1.5, 2.0, \ldots, TA - x - .5$ and $x = BA, \ldots, TA–1$

$$p_{CYIR}(x,a,a,y) = \frac{p_{YIR}(x,a,a,y)}{\sum_{y=1,2,...}^{TA-x-.5} p_{YIR}(x,a,a,y)} \tag{6.3}$$

$$p_{CYIR}(x,i,a,y) = \frac{p_{YIR}(x,i,a,y)}{\sum_{y=1,2,...}^{TA-x-.5} p_{YIR}(x,i,a,y)} \tag{6.4}$$

$y = 1.0, 2.0, \ldots, TA - x - 1$ and $x = BA, \ldots, TA–2$

The expected value $E[CYIR_{x,m,n}]$ measures average time in retirement for those who actually retire. Of course, $E[CYIR_{x,m,n}] > E[YIR_{x,m,n}]$.

Define years-to-retirement *YTR* as *YFS* if the final separation is into inactivity before death. Since final separation could be into death, *YTR* is not defined on these events, and so is a defective random variable. The probability mass function obeys recursions which are very similar to those of *YFS*. In (3.5) the boundary condition $p_{YFS}(x,a,a,.5) = {}^{aa}p_x^d + {}^{aa}p_x^i \, p_{YFS}(x+1,a,i,0)$ becomes the simpler condition that $p_{YTR}(x,a,a,.5) = {}^{aa}p_x^i \, p_{YTR}(x+1,a,i,0)$, and boundary condition (3.6) $p_{YFS}(x,i,a,.5) = {}^{ia}p_x^d + {}^{ia}p_x^i \, p_{YFS}(x+1,a,i,0)$ becomes the simpler condition $p_{YTR}(x,i,a,.5) = {}^{ia}p_x^i \, p_{YTR}(x+1,a,i,0)$. All other recursions are the same for *YTR* and *YFS*.

Death is one way to separate from the labor force while active; but the probability of death does not contribute to *YTR* in the recursions above because retirement requires dying while inactive. The *YTR* random variable is defective in the sense that the sum of its probability mass is less that 1.00 because of the deleted death probabilities from (3.5) and (3.6). We therefore define the conditional-years-to-retirement random variable *CYTR* as years to retirement for those who actually retire (*i.e.*, die while inactive). *CYTR* pmfs are defined in (7.1)-(7.4), where the division cures the defects.

$$p_{CYTR}(x,a,i,y) = \frac{p_{YTR}(x,a,i,y)}{\sum_{y=0,1.5,2.5,...}^{TA-x-.5} p_{YTR}(x,a,i,y)}, \tag{7.1}$$

$$p_{CYTR}(x,i,i,y) = \frac{p_{YTR}(x,i,i,y)}{\sum_{y=0,1.5,2.5,...}^{TA-x-.5} p_{YTR}(x,i,i,y)}, \tag{7.2}$$

for $y = 0, 1.5, 2.5, \ldots, TA - x - .5$ and $x = BA, \ldots, TA–1$

$$p_{CYTR}(x,a,a,y) = \frac{p_{YTR}(x,a,a,y)}{\sum_{y=.5,1.5,...}^{TA-x-.5} p_{YTR}(x,a,a,y)} \tag{7.3}$$

$$p_{CYTR}(x,i,a,y) = \frac{p_{YTR}(x,i,a,y)}{\sum_{y=.5,1.5,...}^{TA-x-.5} p_{YTR}(x,i,a,y)}, \tag{7.4}$$

$y = .5, 1.5, 2.5, \ldots, TA - x - .5$ and $x = BA, \ldots, TA–2$

For a person in labor force state *m* at age *x*–1 and state *n* at age *x*, we define the expected age at retirement as $x + E[CYTR_{x,m.n}]$.

To estimate transition probabilities, data were extracted from the *SIPP* Utilities application developed by Unicon Research Corporation (2013) for the 2001 and 2004 panels. For the 2008 panel, data were extracted using the Data Ferret application provided by the U.S. Census Bureau (2008) and Center for Economic and Policy Research (2012). Individuals are not assigned a one-field unique ID in *SIPP* panels because sampling is based on households; both the household number and person number must be utilized to uniquely identify an individual. To reflect this element of the survey's design, a string variable was created based on the Sample Unit Identifier SSUID (in this context, sample unit refers to household) and person number EPPPNUM to be used as an individual's unique identifier. The string variable was a concatenation of the two fields. This unique ID was used to follow an individual through waves to identify transition states.

To obtain the largest sample size in the context of *SIPP*'s unique design, we used responses based on wave and reference months rather than calendar months. An individual's labor force participation status during the first reference month of the first wave of questioning was used as the initial state of activity or inactivity. This differs from using calendar months in that the calendar date of the first reference month differs for members of various rotation groups. Measurements of second labor force status were taken over 12 months (or three waves) after the date of the initial status measurement for a given individual, and third status was observed 24 months (six waves) later.