Probability mass functions for additional years of labor market activity induced by the Markov (increment–decrement) model

Gary Skoog[a], James Ciecka[b],*

[a]*Department of Economics, DePaul University and Legal Econometrics, Chicago, IL 60604-2287, USA*
[b]*Department of Economics, DePaul University, Chicago, IL 60604-2287, USA*

Abstract

This paper finds probability mass functions for additional years of labor market activity by using recursive formulae. Empirical results are presented for parameters of some mass functions and for probability intervals around those parameters.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Labor force participation; Markov; Probability; Worklife

JEL classification: J21

1. Introduction

This paper introduces the concept of a probability mass function defined for initially active and inactive individuals. The probability mass function is a general concept that enables one to describe some types of specific labor force phenomena as special cases. The specific application in this paper is the creation of an additional-years-of-activity random variable that, among other characteristics, has the property that its expected value is the concept of worklife expectancy. The probability mass function for this random variable allows us to answer questions that heretofore have not been answered. For example, what is the size of an interval that brackets worklife expectancy at a chosen level of probability? What is the shape of the probability distribution describing additional years of labor market activity and how does this distribution change with age? Other than worklife expectancy, what are other measures (like median and mode) of central tendency of years of activity at various ages? The paper begins with a brief discussion of the probability mass function for additional years of

---

*Corresponding author. Tel.: +1-312-362-8831; fax: +1-312-362-5452.

E-mail addresses: gskoog@umich.edu (G. Skoog), jciecka@depaul.edu (J. Ciecka).

life and life expectancy; then we develop the probability mass function for the additional-years-of-activity random variable and describe some empirical results.

2. Probability mass function for additional years of life and life expectancy

Consider a life-mortality problem in which l_x denotes the number of people in a demographic group alive at exact age x. Define T_x as an additional-years-of-life random variable that has probability mass function (pmf)

$$p(x, t) = (l_{x+t} - l_{x+t+1})/l_x \quad t = 0, 1, \ldots, n \tag{1}$$

for $T_x = t$. The curtate life expectancy e_x at age x is

$$e_x = \sum_{t=0}^{n} t(l_{x+t} - l_{x+t+1})/l_x = \sum_{t=0}^{n} tp(x, t) \tag{2a}$$

where age $x + n$ is the youngest age beyond which there are no survivors. In addition to the expectation e_x, the pmf enables one to calculate several other characteristics of additional years of life (e.g. the median, mode, skewness, kurtosis, and various probability intervals).

In (2a) there is an implicit assumption that deaths occur at the beginning of the period between ages $x + t$ and $x + t + 1$. If we think of deaths occurring at the end of the age interval $x + t$ and $x + t + 1$, then life expectancy becomes

$$e_x' = \sum_{t=0}^{n} (t + 1)(l_{x+t} - l_{x+t+1})/l_x. \tag{2b}$$

If we assume that deaths occur uniformly throughout age intervals or at the midpoints of age intervals, then the complete life expectancy is

$$\dot{e}_x = \sum_{t=0}^{n} (t + 0.5)(l_{x+t} - l_{x+t+1})/l_x. \tag{2c}$$

From (2a), (2b), and (2c), it is easily shown that $\dot{e}_x = e_x + 0.5 = e_x' - 0.5$.

Finding the pmf for an additional-years-of-labor-market-activity random variable is, of course, a much more daunting challenge than specifying the additional-years-of-life random variable. In the life-mortality problem, transitions occur in only one direction, i.e. from living to dead. In a Markov process model of labor markets, a person can move from active or inactive status to the death state; but transitions also occur from active to inactive status and from inactive to active status. This two-way movement makes the nature of the problem much more difficult than the life-mortality problem; nevertheless the use of a random variable and its pmf, analogous to T_x and its pmf, are the fundamental concepts.

3. Probability mass functions for additional years of labor market activity

We use the following notation. Let Z_x denote the additional-years-of-labor-market-activity random variable, and let $p(x, s, z)$ denote the probability that a person who is in state s at exact age x will be active for $Z_x = z$ years in the labor force. Let BA (beginning age) and TA (truncation age) denote the ages at which labor force activity commences and ceases, respectively. We consider ages $x = BA, \ldots,$ $TA - 1$, years of activity $z = 0, 1, \ldots, TA - x$, and $s \in \{a, i\}$ where a and i denote active and inactive in the labor force, respectively. The probability that a person who is in state m at age x will be in state n at age $x + 1$ is denoted by $^{m}p_x^{n}$ where $m \in \{a, i\}$, $n \in \{a, i, d\}$, and d is the death state. We follow a Markov, or increment–decrement, model of labor market activity (Bureau of Labor Statistics, 1982, 1986). As in the mortality problem, we could assume beginning, midpoint, or end of period transitions. In this paper we develop $p(x, s, z)$ with an end-of-period assumption and present some empirical results under that assumption, but we also state the probability mass functions with beginning and midpoint transitions (see Skoog, 2002).

Probability mass functions are defined by the combination of global conditions (that hold regardless of whether transitions occur at the beginning, mid-point, or end of a period), boundary conditions (that hold for 0, 0.5, and 1 year of activity depending on when transitions occur), and main recursions (that are interior to the ages treated by the boundary conditions). The global conditions are stated in (3a)–(3c).

Global conditions

$$p(x, a, r) = p(x, i, r) = 0 \quad \text{if } r < 0 \text{ or } r > TA - x \tag{3a}$$

$$p(TA, a, 0) = p(TA, i, 0) = 1 \tag{3b}$$

$$^{a}p_x^{a} = {}^{i}p_x^{a} = 0 \quad \text{for } x \geq TA. \tag{3c}$$

Condition (3a) merely expresses the idea that there is a zero probability of accumulating negative years of activity or years of activity in excess of the number of years to the truncation age. (3b) states that the probability is one that a person age $x = TA$ accumulates zero years of activity. Since (3c) says that labor force activity ceases at age $x = TA$, it also implies (3b) and illustrates the idea that we seek boundary and global conditions on the mass functions with, or implied by, conditions on the underlying transition probabilities.

The pmf with end of period transitions is stated and explained below.

Probability mass function for $Z = z$ years of activity with end of period transitions. Boundary conditions:

$$p(x, a, 0) = 0 \quad x = BA, \ldots, TA - 1 \tag{4a}$$

$$p(TA - 1, i, 0) = 1 \tag{4b}$$

$$p(x, i, 0) = {}^{i}p_x^{i}\, p(x + 1, i, 0) + {}^{i}p_x^{d} \quad x = BA, \ldots, TA - 2 \tag{4c}$$

$$p(x, a, 1) = {}^{a}p_x^{i}\, p(x + 1, i, 0) + {}^{a}p_x^{d} \quad x = BA, \ldots, TA - 1. \tag{4d}$$

In (4a) and (4b) we observe that an active person at age x must accumulate at least 1 year of activity because transitions occur at the end of the age interval $(x, x + 1)$, implying that $p(x, a, 0) = 0$; and an inactive person age $TA - 1$ will, with certainty, have zero years of activity, thus $p(TA - 1, i, 0) = 1$. Condition (4c) expresses the observation that there are only two ways a person inactive at age x can experience no additional years of activity: die or remain inactive a year and repeat the process. In (4d) a person active at age x accumulates exactly 1 year of future activity by turning inactive or through death.

The remainder of the probability mass values are defined by the main recursions in (4e) and (4f) for ages $x = BA, \ldots, TA - 1$.

Main recursions

$$p(x, a, z) = {}^{a}p_{x}^{a} p(x + 1, a, z - 1) + {}^{a}p_{x}^{i} p(x + 1, i, z - 1) \quad \text{for } z = 2, 3, \ldots, TA - x \tag{4e}$$

$$p(x, i, z) = {}^{i}p_{x}^{a} p(x + 1, a, z) + {}^{i}p_{x}^{i} p(x + 1, i, z) \quad \text{for } z = 1, 2, \ldots, TA - x. \tag{4f}$$

The right-hand side of (4e) is the sum of two terms that contribute to the probability that an active person age x will accumulate z years of activity: $p(x + 1, a, z - 1)$ and $p(x + 1, i, z - 1)$ are the probabilities of experiencing $z - 1$ active years when being active and inactive at age $x + 1$, respectively; the probability of z years of activity results when the former probability is multiplied by ${}^{a}p_{x}^{a}$ and the latter by ${}^{a}p_{x}^{i}$. In (4f), ${}^{i}p_{x}^{a}$ and ${}^{i}p_{x}^{i}$ are probabilities for people who start age x as inactive and thereby cannot experience any activity over the interval $(x, x + 1)$; therefore, in order to accumulate z years of activity, these probabilities must be multiplied by probabilities $p(x + 1, a, z)$ and $p(x + 1, i, z)$ for people age $x + 1$ who have already accumulated z active years.

The recursions that define the probability mass functions when transitions occur at the beginning and at the midpoint of age intervals are stated in (5a)–(5d) and (6a)–(6e). We use Y to denote the years-of-activity random variable with transitions at the beginning of age intervals and W when transitions occur midyear.

Probability mass function for $Y = y$ years of activity with beginning of period transitions.

Boundary conditions:

$$p(x, a, 0) = {}^{a}p_{x}^{i} p(x + 1, i, 0) + {}^{a}p_{x}^{d} \quad x = BA, \ldots, TA \tag{5a}$$

$$p(x, i, 0) = {}^{i}p_{x}^{i} p(x + 1, i, 0) + {}^{i}p_{x}^{d} \quad x = BA, \ldots, TA. \tag{5b}$$

Main recursions:

$$p(x, a, y) = {}^{a}p_{x}^{a} p(x + 1, a, y - 1) + {}^{a}p_{x}^{i} p(x + 1, i, y) \tag{5c}$$

$$p(x, i, y) = {}^{i}p_{x}^{a} p(x + 1, a, y - 1) + {}^{i}p_{x}^{i} p(x + 1, i, y). \tag{5d}$$

In (5c) and (5d), $x = BA, \ldots, TA - 1$ and $y = 1, 2, \ldots, TA - x$.

Probability mass function $W = w$ years of activity with midpoint transitions.

Boundary conditions

$$p(x, a, 0.5) = {}^a p_x^d + {}^a p_x^i p(x + 1, i, 0) \quad x = BA, \ldots, TA - 1 \tag{6a}$$

$$p(x, a, 0) = 0 \quad x = BA, \ldots, TA - 1 \tag{6b}$$

$$p(x, i, 0) = {}^i p_x^d + {}^i p_x^i p(x + 1, i, 0) \quad x = BA, \ldots, TA - 1. \tag{6c}$$

Main recursions

$$p(x, a, w) = {}^a p_x^a p(x + 1, a, w - 1) + {}^a p_x^i p(x + 1, i, w - 0.5) \tag{6d}$$

$$p(x, i, w) = {}^i p_x^a p(x + 1, a, w - 0.5) + {}^i p_x^i p(x + 1, i, w). \tag{6e}$$

In (6d) and (6e), $x = BA, \ldots, TA - 1$; $w = 1, 1.5, 2, \ldots, TA - x$ in (6d); and $w = 0.5, 1.0, 1.5, \ldots,$ $TA - x$ in (6e).

The probability mass function for Z_x can be used to define the expected value of additional years of activity (i.e. worklife expectancy) for actives at age x as

$$E(Z_x) = WLE_x(a) = \sum_{z=0}^{TA-x} zp(x, a, z) = {}^a e_x^a \tag{7a}$$

and worklife for inactives as

$$WLE_x(i) = \sum_{z=0}^{TA-x} zp(x, i, z) = {}^i e_x^a. \tag{7b}$$

In addition to $E(Z_x)$ the probability mass function for Z_x can be used to define other properties of labor market activity that were heretofore impossible to calculate. For example, the standard deviation, median, mode, skewness, and kurtosis can now be calculated.[1] Using transition probabilities ${}^m p_x^n$ for 1997–1998 (Ciecka et al., 2000), we have calculated the pmf for $BA = 16$ through $TA = 80$. Summary statistics for these pmf are presented in Table 1 for men at ages 20, 35, 50, and 65 and Fig. 1 shows their graphs. The pmf are skewed to the left for ages 16–48, and skewed to the right for ages 49–77.

Table 1
Properties of years-of-activity random variable at various ages for initially active men

Property	Age 20	Age 35	Age 50	Age 65
Mean (WLE)	37.71	25.46	13.05	4.61
Median	38.72	25.85	12.53	3.41
Mode	41.00	27.00	13.00	1.00
Standard deviation	9.35	7.55	5.48	3.15
Skewness	−1.15	−0.63	0.11	0.98
Kurtosis	5.01	3.56	2.78	3.47
50% Probability interval around mean	(36.13, 45.00)	(23.10, 31.00)	(10.00, 16.37)	(1.84, 4.61)
50% Probability interval around median	(36.13, 45.00)	(23.10, 31.00)	(10.00, 16.37)	(1.00, 3.41)
50% Probability interval around mode	(36.13, 45.00)	(23.10, 31.00)	(10.00, 16.37)	(1.00, 3.41)

---

[1] Throughout the rest of this paper, all results are for active men. Related results could be stated for women and inactives.

PMF for Active Men Age 20

PMF for Active Men Age 35

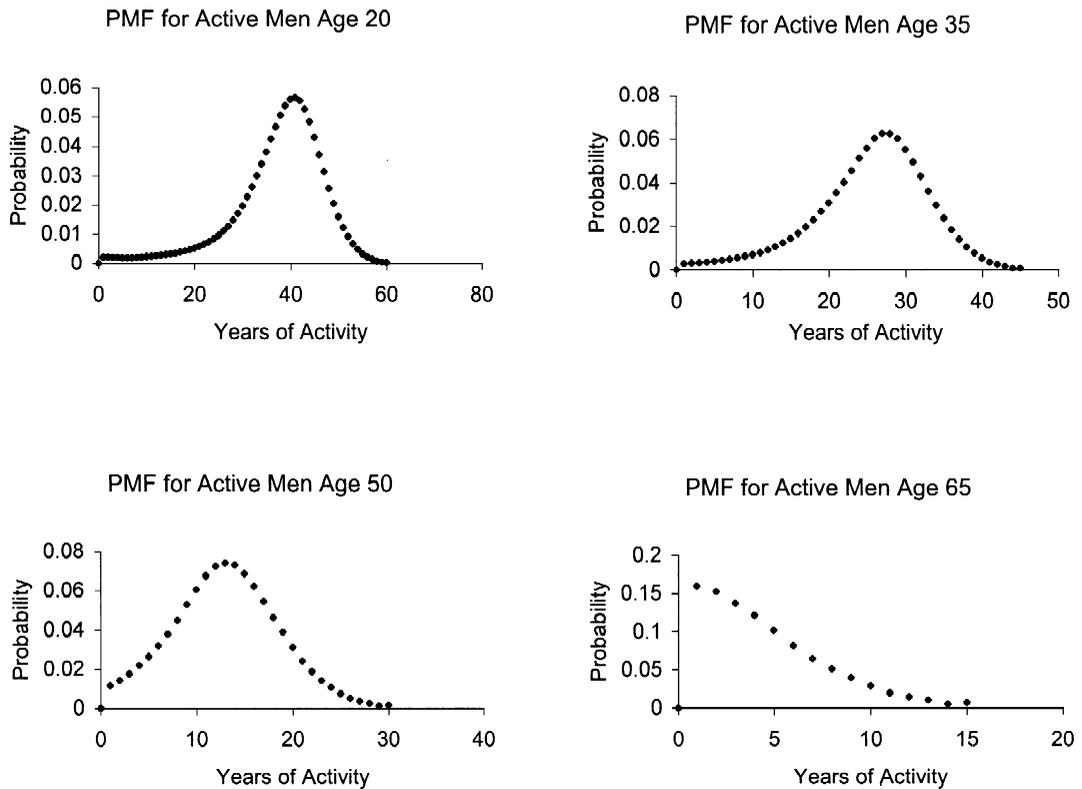PMF for Active Men Age 50

PMF for Active Men Age 65

Fig. 1. Probability mass functions for initially active men at ages 20, 35, 50 and 65.

The skewness parameter is small (about 0.3 or less) between ages 43 and 53. It is in the age range 43–53 that kurtosis is ~3 as would be the case for normally distributed random variables; and means, medians and modes are approximately equal to each other (e.g. they are 18.71, 18.64 and 20, respectively, at age 43 and 10.79, 10.06, and 10.00 at age 53). With the pmf in hand, we also can calculate probability intervals that include various parameters. One possible use of such intervals is in the area of forensic economics. Rather than construct the usual ±1.96 standard error estimates around worklife expectancy as the US Supreme Court sanctioned in Hazelwood (1977) and Casteneda (1977), it would be more accurate to compute probability intervals (PI) directly from the distribution rather than from a normal approximation. For example, forensic economists may be interested in the smallest[2] 50% interval around the mean of Z_x, where 50% is chosen to correspond to the legal concept of a statement or calculation 'being accurate to within a reasonable degree of economic certainty'. Alternatively, such intervals may be interpreted as characterizing the legal concept of 'more likely true than not true'. For example, at age 20 the 50% PI is (36.13, 45.00) for the mean, median, and mode. The mode is 41 years, and the PI contains 4.87 years to the left of the mode and 4.00 years to the right. Since the mean is 37.71 years, there are only 1.58 years to the left of the mean and 7.29

---

[2]Other probability intervals, such as those allocating equal probability in each tail, may be of interest and can easily be computed.

years to the right. Therefore, an outcome at falls in the 50% PI is more likely to exceed the mean (i.e. worklife expectancy) than to be smaller than the mean. At age 50 the pmf is nearly symmetrical since its mean, median, and mode are all close to each other (13.05, 12.53, and 13.00, respectively).

3. Conclusion

This paper proffers the concept of a probability mass function for years of additional activity and presents empirical results for men defined conditionally on initial active status with end of period transitions. Similar results could have been presented for initially inactive men, for initially active or inactive women, for various levels of education for either men or women, and for beginning or mid-period transitions. By defining additional years of labor market activity as a random variable, we are able to make probabilistic statements about parameters such as *WLE* and completely describe the probability distribution of years of activity. We can calculate probability intervals, place bounds of error, and make probabilistic statements that were not possible previously. In addition, the idea of a pmf could be applied to different types of random variables beyond additional years of activity. Such random variables include the number of years until final separation from the labor force, number of lifetime labor force separations and accessions, number of years to retirement, and years spent in retirement.

References

Casteneda v. Partida, 1977. 430 US 482.

Bureau of Labor Statistics, 1982. Tables of Working Life: The Increment–decrement Model, Bulletin 2135. US Government Printing Office, Washington, DC.

Bureau of Labor Statistics, 1986. Worklife Estimates: The Effects of Race and Education, Bulletin 2254. US Government Printing Office, Washington, DC.

Ciecka, J., Donley, T., Goldman, J., 2000. A Markov process model of worklife expectancies based on labor market activity in 1997–98. Journal of Legal Economics 9 (3), 33–68.

Hazelwood School District v. United States, 1977. 433 US 299.

Skoog, G., 2002. Worklife expectancy: theoretical results. Paper presented at ASSA Meeting, Atlanta.