Gary R. Skoog and James E. Ciecka. 2012. An Autoregressive Model of Order Two for Worklife Expectancies and Other Labor Force Characteristics with an Application to Major League Baseball Hitters. *Journal of Legal Economics* 18(2): pp. 47–78.

An Autoregressive Model of Order Two for Worklife Expectancies and Other Labor Force Characteristics with an Application to Major League Baseball Hitters*

Gary R. Skoog and James E. Ciecka

I. Introduction

The Markov process model has been the standard for worklife related research since 1982 when the Bureau of Labor Statistics introduced its Increment-Decrement (Markov model) construct in *Bulletin 2135* and updated its worklife estimates in *Bulletin 2254* (1986). Skoog and Ciecka (2001a, 2001b, and 2002) extended Bureau of Labor Statistics methodology by specifying recursions which captured entire probability distributions implicit in the Markov model. However, a Markov model, by definition, remains an autoregressive process of order one (AR(1)). Being active in the labor force at age x implies certain transition probabilities to remaining active or moving to inactive at age $x+1$ regardless of a person's active and inactive states at ages $x-1$, $x-2, \ldots$. Similarly, inactivity at age x affects activity status at age $x+1$ in a manner that does not depend on status at ages younger than age x. This paper deals with an autoregressive model of order two (AR(2)).[1] That is, a person's labor force states at ages $x-1$ and age x influence status at age $x+1$. We take one step backwards in time and incorporate information at time x and $x-1$ into transition probabilities. We specify recursions which define

Gary R. Skoog, DePaul University, Department of Economics, and Legal Econometrics, Inc., 1527 Basswood Circle, Glenview IL 60025-2006. Tel: 847-729-6154; Fax: 847-729-6158; Email: gskoog@umich.edu. James E. Ciecka, Department of Economics, DePaul University, 1 East Jackson Blvd., Chicago IL 60604. Telephone: 312 362-8831; Fax: 312 362-5452; Email: jciecka@depaul.edu

* The authors thank our ASSA discussant Edward Foster and referees. Their comments and insights could only be the result of a careful line-by-line reading of our paper. We thank them for the time and energy that task entailed. Their efforts led to a much improved and more readable paper. We also thank Alex Krautmann for compiling data used in Section III.

probability mass functions (pmfs) for future years of activity for four combinations of states at ages $x - 1$ and age x: (active, active), (active, inactive), (inactive, active), and (inactive, inactive). These pmfs enable us to calculate various measures of central tendency (mean, *i.e.*, worklife expectancy, median, and mode), spread and shape (standard deviation, skewness, and kurtosis), and probability intervals (inter-quartile range and 90% interval). We exemplify our AR(2) recursions and pmfs with an application to major league baseball hitters. With the proper theory in place, similar applications only await a sufficiently large data set tracking labor force activity of individuals by gender and education simultaneously at ages $x - 1$, x, and $x + 1$ as opposed to only ages x and $x + 1$ in a Markov model.

The specific application of the AR(2) model in this paper will be of only limited value unless a forensic economist happens to be dealing with a major league baseball hitter. However, this specific application suggests that, whether initially active for two years in a row within the AR(2) model or initially active for just one year in the Markov (AR(1)) model, worklife expectancies produced by the AR(1) and AR(2) models are in close agreement with each other. On the other hand, initial inactivity for two years within the AR(2) model leads to much smaller worklives than initial inactivity within the Markov model. However, these results are not the main intended contribution of the paper because we do not know whether they can be extended to the labor force disaggregated by both gender and education – the standard disaggregation achieved with the Markov model. Rather, the main contribution of this paper is the specification of the set of recursive formulae needed to calculate probability mass functions for years of activity within the AR(2) model without having to resort to approximations through simulations. The other main contribution of this paper is an analysis of the relationships between AR(1) and AR(2) models and other results like worklife expectancy recursions in an AR(2) context. We show which transition probability inequalities within the AR(2) and AR(1) models must carry over to their worklife expectancies, and show that transition inequalities across AR(2) and AR(1) models do not carry over generally to worklife expectancies.

This paper is organized as follows. Section II contains notation and recursions used to calculate probability mass functions within the AR(2) model. The results of this section of the paper can be applied to all data sets containing the requisite transition probabilities. Section III contains an application to baseball. Section IV contains a general analysis of the AR(2) model and relationships between it and the Markov model. Section V is a conclusion.

II. Notation and Recursions for AR(2) Model

We use the following notation: a denotes active in the labor force, i denotes inactive, and d the death state. Let $k \in \{a, i\}$, $m \in \{a, i\}$, and $n \in \{a, i, d\}$. A person's exact current age is denoted by x with BA being the youngest age at which labor market activity can occur. Everyone alive must have died by truncation age TA. ${}^{km}p_x^n$ denotes the probability that a person in state k at age $x - 1$ and state m at age x will be in state n at age $x + 1$. Transitions between states are assumed to occur at the midpoint between ages; anyone alive, whether in state a or i, at age $TA - 1$ transitions to state d at age $TA - .5$. $YA_{x,k,m}$ denotes a random variable of future active time for a person age x who was in state k at age $x - 1$ and state m at age x, and $p_a(x, k, m, y)$ measures the probability that a person age x who was in state k at age $x - 1$ and state m at age x will accumulate y additional years of activity. That is, $p_a(x, k, m, y)$ measures the probability that $YA_{x,k,m} = y$. Recursions defining $p_a(x, k, m, y)$ consist of global conditions (GC), boundary conditions (BC), and main recursions (MR). GC refers to extreme values of y and x as well as conditions that hold at all ages. BC deals with probabilities of future labor force activity being either 0 or .5 years. MR captures probabilities of future activity for years y exceeding values defined in BC.[2] See formulae GC1-GC4, BC1-BC5, and MR1-MR4 below.

GC1 says that future years of activity cannot be negative nor can they exceed $TA - x - .5$ years. The latter condition holds because everyone alive at age x dies before or at age $TA - .5$. Since everyone has died by age TA, the probability of zero future labor force time at age TA is 1.00 as expressed in GC2. Global condition GC3 expresses the assumption that transition to the death state is independent of previous labor force states. The last global condition, GC4, says that whatever a person's states at ages $x - 1$ and x, a transition must occur to either the a, i, or d state.

Condition BC1 expresses the impossibility of no future labor force time if a person were active at age x. However, an inactive person at age x, whether active or inactive at age $x - 1$, can accumulate no additional labor force time by transitioning to d or remaining in i thereafter (conditions BC2 and BC3). At a minimum, at least .5 years of activity must follow if a person were active at age x, whether active or inactive at age $x - 1$; BC4 and BC5 show how this minimal amount of activity would occur, *i.e.*, by transitioning to the d state or to the i state and remaining inactive thereafter. We note that BC2 and BC3 resemble each other but yield different probabilities since the former depends on ${}^{ai}p_x^i$ while the latter depends on ${}^{ii}p_x^i$. In a similar manner, BC4 and BC5 are functions of ${}^{aa}p_x^i$ and ${}^{ia}p_x^i$, respectively.

The remaining probability mass values are defined by the main recursions. The right-hand side of MR1 is the sum of two terms that contribute to the probability that an active person ages $x - 1$ and x will accumulate y years of activity: (1) The first term ${}^{aa}p_x^a p_a(x + 1, a, a, y - 1)$ is the product of two factors. The second factor, $p_a(x + 1, a, a, y - 1)$, is the probability that a person active at ages x and $x + 1$ will have $y - 1$ years of future activity and, when multiplied by ${}^{aa}p_x^a$ yields another year of activity at age x. (2) The second term ${}^{aa}p_x^i p_a(x + 1, a, i, y - .5)$ also is the product of two factors. The second factor $p_a(x + 1, a, i, y - .5)$ is the probability that a person active at age x and inactive at age $x + 1$ will have $y - .5$ years of future activity and, when multiplied by ${}^{aa}p_x^i$ yields an additional .5 year of activity at age x. The second factors in both terms aggregate sample paths resulting from remaining active for $y - 1$ and $y - .5$ years, respectively, from age x+1; and their respective multipliers ${}^{aa}p_x^a$ and ${}^{aa}p_x^i$ induce an additional one whole year and .5 of a year of activity. The remaining main recursions work in a manner similar to MR1. In each case, the left transition probability superscripts (on the right side of a recursion) denote activity states that always match the states at ages $x - 1$ and x on the left side of the recursion; and the second left and right transition probability superscripts match activity states at ages x and $x + 1$ on the right side of a recursion. For example, the left transition probability superscripts *ia* for ages $x - 1$ and x match the inactivity/activity arguments in $p_a(x, i, a, y)$ on the left side of recursion MR2. In addition, the second left and right transition probability superscripts *aa* for ages x and $x + 1$ in the first term on the right side of MR2 match the probability activity/activity arguments in $p_a(x + 1, a, a, y - 1)$; and the second left and right transition probability superscripts *ai* in the second term on the right side of MR2 match the probability activity/inactivity arguments in $p_a(x + 1, a, i, y - .5)$.

Our goal was to find the exact probability distribution of the years of activity random variable $Y_{x,k,m}$ within the context of the AR(2) model. Recursions GC1-GC4, BC1-BC5, and MR1-MR4 not only accomplish that task, but they do so in a computationally efficient manner. Once transition probabilities have been estimated, these recursions yield the exact distribution of $Y_{x,k,m}$ in a few seconds of computer time for all $x = 16, \ldots, 111$. The computational efficiency itself is remarkable when one considers the number of sample paths of activity, inactivity, and death that could occur; and each path makes its own contribution to the pmf for $Y_{x,k,m}$. In general, $2^{TA-x} - 1$ sample paths can occur – a number impossible to compute and store if $TA - x$ were as small as (say) 50. Such an unwieldy number of paths occurs because of the possibility of back and forth transitions between activity and inactivity. Here, a comparison to the additional-years-of-life random variable, YL_x, may be useful. That random variable, whose mean is life expectancy, measures movement in only one direction, *i.e.*, from living to living or living to dead but, of course, not from dead to living. The resulting number of sample paths is only $TA - x$ at age x. Although a convenient recursion exists for YL_x, its distribution could be found by simply following all sample paths and computing their associated probabilities. Therefore, a recursion for YL_x, although simple and elegant, is unnecessary. On the other hand, recursions for $Y_{x,k,m}$ are indispensable because there simply are too many paths to follow, compute their probabilities, and store results.

Recursions Defining Probability Mass Functions for $YA_{x,k,m}=y, k \in \{a, i\}, m \in \{a, i\}$ for AR(2) Model with Midpoint Transitions

Global Conditions

GC1 $\quad p_a(x,a,a,y) = p_a(x,a,i,y) = p_a(x,i,a,y) = p_a(x,i,i,y) = 0$

for $y < 0$ or $y > TA - x - .5$

GC2 $$p_a(TA,a,a,0) = p_a(TA,a,i,0) = p_a(TA,i,a,0) = p_a(TA,i,i,0) = 1$$

GC3 $\quad {}^{aa}p_x^d = {}^{ai}p_x^d = {}^{ia}p_x^d = {}^{ii}p_x^d = {}^{\bullet}p_x^d$

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

GC4 $\quad {}^{aa}p_x^a + {}^{aa}p_x^i + {}^{\bullet}p_x^d = 1; \quad {}^{ai}p_x^a + {}^{ai}p_x^i + {}^{\bullet}p_x^d = 1;$

$\quad {}^{ia}p_x^a + {}^{ia}p_x^i + {}^{\bullet}p_x^d = 1; \quad {}^{ii}p_x^a + {}^{ii}p_x^i + {}^{\bullet}p_x^d = 1$

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

Boundary Conditions

BC1 $\quad p_a(x,a,a,0) = p_a(x,i,a,0) = 0$

BC2 $\quad p_a(x,a,i,0) = {}^{ai}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$

BC3 $\quad p_a(x,i,i,0) = {}^{ii}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$

BC4 $\quad p_a(x,a,a,.5) = {}^{aa}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$

BC5 $\quad p_a(x,i,a,.5) = {}^{ia}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$

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

Main Recursions

MR1 $$p_a(x,a,a,y) = {}^{aa}p_x^a p_a(x+1,a,a,y-1) + {}^{aa}p_x^i p_a(x+1,a,i,y-.5)$$

MR2 $$p_a(x,i,a,y) = {}^{ia}p_x^a p_a(x+1,a,a,y-1) + {}^{ia}p_x^i p_a(x+1,a,i,y-.5)$$

for $y = 1.5, 2.5, 3.5, \ldots, TA - x - .5$

MR3 $$p_a(x,a,i,y) = {}^{ai}p_x^a p_a(x+1,i,a,y-.5) + {}^{ai}p_x^i p_a(x+1,i,i,y)$$

MR4 $$p_a(x,i,i,y) = {}^{ii}p_x^a p_a(x+1,i,a,y-.5) + {}^{ii}p_x^i p_a(x+1,i,i,y)$$

for $y = 1, 2, 3, \ldots, TA - x - 1$

We calculate the characteristics of $Y_{x,k,m}$ with the following formulae:

The expected value of $Y_{x,k,m}$, worklife expectancy, is defined by

$$E(Y_{x,k,m}) = WLE_x(k,m) = \sum_y y p_a(x,k,m,y) = {}^{km}e_x^a.$$

The median value of $Y_{x,k,m}$, y_{med}, possess the property such that

$$\Pr(Y_{x,k,m} \le y_{med}) \ge .50 \quad \text{and} \quad \Pr(Y_{x,k,m} \ge y_{med}) \ge .50$$

and the mode of $Y_{x,k,m}$, y_{mode}, is the value of $Y_{x,k,m}$ that

$$p_a(x,k,m,y_{mode}) \ge p_a(x,k,m,y)$$

for all values of y.

The variance, standard deviation, skewness, and kurtosis are defined by

$$V(Y_{x,k,m}) = \sum_y (y - {}^{km}e_x^a)^2 p_a(x,k,m,y),$$

$$SD(Y_{x,k,m}) = \sqrt{V(Y_{x,k,m})},$$

$$SK(Y_{x,k,m}) = (1/SD(Y_{x,k,m}))^3 \sum_y (y - {}^{km}e_x^a)^3 p_a(x,k,m,y),$$

and

$$KU(Y_{x,k,m}) = (1/SD(Y_{x,k,m}))^4 \sum_y (y - {}^{km}e_x^a)^4 p_a(x,k,m,y).$$

Cumulative probabilities occur at values of $Y_{x,k,m}$ where

$$\Pr(Y_{x,k,m} \leq y_\alpha) \geq \alpha \quad \text{and} \quad \Pr(Y_{x,k,m} \geq y_\alpha) \geq 1 - \alpha$$

for $\alpha = .10, .25, .75, .90$.

Two modified sets of recursions deal with transitions between states occurring at the beginning and end, rather than the midpoint, of periods. Global conditions remain unchanged from midpoint transitions and are not repeated. However, boundary and main recursions do change as indicated below. With beginning-of-year transitions (see recursions marked with asterisks), the fundamental idea involves recognizing a gain of one year of activity when a transition occurs between *aa* or *ia* but no increase for an *ai* or *ii* transition. Under the assumption of end-of-year transitions (see recursions marked with double asterisks), a gain of one year of activity occurs with an *aa* or *ai* transition but no increase for an *ia* or *ii* transition. Beginning, middle, and ending transitions affect worklife expectancies in the following manner: (1) With beginning-of-period transitions and being active at age x, worklife at age x is one-half year shorter than under the assumption of mid-year transitions. (2) Assuming end-of-year transitions and starting active, worklife at age x is one-half year longer than under the assumption of mid-year transitions. (3) For those inactive at age x, worklives are identical under all three transition point assumptions. (4) The pmf (when starting active) based on beginning-of-period transitions is the same as the pmf with mid-period transitions except the former is shifted one-half unit to the left of the latter. Similarly, the pmf based of end-of-period transitions is one-half unit to the right of its mid-period transitions counterpart. Therefore, the mean and all percentile points increase one-half unit as we move from beginning to mid-point to ending transitions; but the variance, standard deviation, skewness, and kurtosis remain unchanged. (5) When starting inactive, pmfs for activity do not depend on time of transitions; therefore all measures of activity are the same under all three timing assumptions. (6) Finally, we note that mid-year transitions seem appropriate in most labor market settings as in a life or survivor table; but beginning or end-of-year transitions may be appropriate in some situations, *e.g.*, the application presented in the next section utilizes end-of-period transitions.

Recursions Defining Probability Mass Functions for $YA_{x,k,m}=y, k \in \{a, i\}$, $m \in \{a, i\}$ for AR(2) Model with Beginning-of-Year Transitions
Boundary Conditions

BC1* $\quad p_a(x,a,a,0) = {}^{aa}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$

BC2* $p_a(x,i,a,0) = {}^{ia}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$

BC3* $p_a(x,i,i,0) = {}^{ii}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$

BC4* $p_a(x,a,i,0) = {}^{ai}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$
for $x = BA,\ldots,TA - 1$
Main Recursions

MR1* $p_a(x,a,a,y) = {}^{aa}p_x^a p_a(x+1,a,a,y-1) + {}^{aa}p_x^i p_a(x+1,a,i,y)$

MR2* $p_a(x,i,a,y) = {}^{ia}p_x^a p_a(x+1,a,a,y-1) + {}^{ia}p_x^i p_a(x+1,a,i,y)$
for $y = 1,2,3,\ldots,TA - x - 1$

MR3* $p_a(x,a,i,y) = {}^{ai}p_x^a p_a(x+1,i,a,y-1) + {}^{ai}p_x^i p_a(x+1,i,i,y)$

MR4* $p_a(x,i,i,y) = {}^{ii}p_x^a p_a(x+1,i,a,y-1) + {}^{ii}p_x^i p_a(x+1,i,i,y)$
for $y = 1,2,3,\ldots,TA - x - 1$

Recursions Defining Probability Mass Functions for $YA_{x,k,m}=y$, $k \in \{a, i\}$, $m \in \{a, i\}$ *for AR(2) Model with End-of-Year Transitions*
Boundary Conditions

BC1** $p_a(x,a,a,0) = 0$

BC2** $p_a(x,i,a,0) = 0$

BC3** $p_a(x,i,i,0) = {}^{ii}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$

BC4** $p_a(x,a,i,0) = {}^{ai}p_x^i p_a(x+1,i,i,0) + {}^{\bullet}p_x^d$

BC5** $p_a(x,a,a,1) = {}^{aa}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$

BC6** $p_a(x,i,a,1) = {}^{ia}p_x^i p_a(x+1,a,i,0) + {}^{\bullet}p_x^d$
for $x = BA,\ldots,TA - 1$
Main Recursions

MR1**
$$p_a(x,a,a,y) = {}^{aa}p_x^a p_a(x+1,a,a,y-1) + {}^{aa}p_x^i p_a(x+1,a,i,y-1)$$

MR2**
$$p_a(x,i,a,y) = {}^{ia}p_x^a p_a(x+1,a,a,y-1) + {}^{ia}p_x^i p_a(x+1,a,i,y-1)$$
for $y = 2,3,4.,\ldots,TA - x$

MR3** $p_a(x,a,i,y) = {}^{ai}p_x^a p_a(x+1,i,a,y) + {}^{ai}p_x^i p_a(x+1,i,i,y)$

MR4** $p_a(x,i,i,y) = {}^{ii}p_x^a p_a(x+1,i,a,y) + {}^{ii}p_x^i p_a(x+1,i,i,y)$
for $y = 1,2,3,\ldots,TA - x - 1$

III. Application to Major League Baseball

Pmfs and Their Characteristics for Major League Baseball Hitters

This paper's main contribution consists of specifying recursions for pmfs within the context of an autoregressive model of order two and the specification error analysis in Section IV. We would like to apply our recursions to national labor force data; but a large nationally representative data set like the CPS may not exist that would enable us to trace individuals by age, gender, and education simultaneously long enough to record activity status at ages $x - 1$, x, and $x + 1$. In searching for a data set to apply our recursions, we decided to make an application to major league baseball players; but the methods are completely general. Comprehensive year-by-year records exist for all hitters and pitchers for the entire history since 1871 of major league baseball. In the application that follows, we confine ourselves to hitters in the eleven-year period 1997 – 2007, during which time 1,536 hitters played major league baseball.[3] Although our data set contains 16,896 (= 1,536 players x 11 seasons) maximum potential observations, we used only 8,551 observations because some players entered major league baseball after 1997 and therefore had no record between 1997 and the date they entered the major leagues. Also, we deleted players after they were inactive for four consecutive seasons in order to limit their influence on our estimate of ${}^{ii}p_x^i$ after their careers had effectively ended. A hitter is defined to be an active major league player in a particular year if he appeared in at least one major league game during that year. A hitter is inactive in a particular year if he did not play in any major league games after commencing his major league career. The end-of-period transitions assumption seems most appropriate given our definition that an active player receives credit for a year of activity if he appears in one major league game during a year, *i.e*, once active, the player cannot become inactive until the following season.

Table 1 contains transition probabilities ${}^{km}p_x^n$, $k \in \{a, i\}$, $m \in \{a, i\}$, and $n \in \{a, i, d\}$ for $x = 23, 24, \ldots, 45$. Table 1 also contains the result of estimating first order Markov model transition probabilities ${}^{m}p_x^n$ and comparisons to estimated ${}^{km}p_x^n$ parameters; further discussion of the AR(2)-AR(1) comparison appears in Section IV. Typically we notice that[4]

Table 1. Transition Probabilities for Major League Baseball Hitters with AR(2) and AR(1) Models*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Age	${}^{aa}p_x^a$	${}^{aa}p_x^i$	${}^{ia}p_x^a$	${}^{ia}p_x^i$	${}^{ai}p_x^a$	${}^{ai}p_x^i$	${}^{ii}p_x^a$	${}^{ii}p_x^i$	${}^{a}p_x^a$	${}^{a}p_x^i$	${}^{i}p_x^a$	${}^{i}p_x^i$
23	0.8923	0.1062	0.8129	0.1856	0.5377	0.4609	0.0000	0.9985	0.8506	0.1479	0.4510	0.5476
24	0.8902	0.1084	0.7268	0.2718	0.4312	0.5674	0.3073	0.6913	0.8282	0.1704	0.4029	0.5956
25	0.9028	0.0958	0.6727	0.3259	0.3779	0.6208	0.1332	0.8655	0.8349	0.1637	0.3073	0.6914
26	0.8535	0.1452	0.5885	0.4102	0.2903	0.7084	0.0881	0.9106	0.7954	0.2033	0.2010	0.7976
27	0.8567	0.1420	0.5826	0.4161	0.2497	0.7490	0.0679	0.9308	0.8093	0.1894	0.1592	0.8395
28	0.8730	0.1257	0.4577	0.5410	0.1952	0.8035	0.0444	0.9543	0.8113	0.1874	0.1035	0.8952
29	0.8377	0.1610	0.3484	0.6503	0.1347	0.8640	0.0964	0.9022	0.7899	0.2088	0.1110	0.8877
30	0.8531	0.1455	0.4645	0.5342	0.2289	0.7698	0.0591	0.9395	0.8106	0.1880	0.1248	0.8738
31	0.8639	0.1347	0.4112	0.5874	0.1969	0.8017	0.0636	0.9350	0.8184	0.1802	0.1051	0.8935
32	0.8350	0.1635	0.4369	0.5617	0.2031	0.7954	0.0482	0.9503	0.7951	0.2035	0.0930	0.9055
33	0.7919	0.2066	0.4992	0.4992	0.1610	0.8374	0.0086	0.9899	0.7691	0.2293	0.0617	0.9368
34	0.7935	0.2049	0.5376	0.4608	0.0832	0.9152	0.0188	0.9795	0.7797	0.2187	0.0421	0.9563
35	0.7637	0.2346	0.4278	0.5704	0.0565	0.9418	0.0000	0.9983	0.7513	0.2469	0.0177	0.9806
36	0.6938	0.3044	0.3850	0.6132	0.1109	0.8873	0.0000	0.9982	0.6841	0.3141	0.0294	0.9688
37	0.6800	0.3181	0.3465	0.6516	0.0475	0.9505	0.0000	0.9980	0.6618	0.3362	0.0129	0.9851
38	0.6277	0.3702	0.3118	0.6861	0.0688	0.9290	0.0000	0.9978	0.6335	0.3643	0.0152	0.9826
39	0.6176	0.3801	0.2805	0.7171	0.0454	0.9523	0.0000	0.9976	0.6122	0.3855	0.0097	0.9880

(continued on next page)

Table 1. (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Age	${}^{aa}p_x^a$	${}^{aa}p_x^i$	${}^{ia}p_x^a$	${}^{ia}p_x^i$	${}^{ai}p_x^a$	${}^{ai}p_x^i$	${}^{ii}p_x^a$	${}^{ii}p_x^i$	${}^{a}p_x^a$	${}^{a}p_x^i$	${}^{i}p_x^a$	${}^{i}p_x^i$
40	0.5586	0.4389	0.2524	0.7450	0.0713	0.9262	0.0000	0.9974	0.5371	0.4604	0.0131	0.9843
41	0.5439	0.4533	0.2271	0.7701	0.0767	0.9205	0.0000	0.9972	0.5817	0.4155	0.0188	0.9784
42	0.3988	0.5982	0.2044	0.7926	0.0000	0.9970	0.0000	0.9970	0.4985	0.4985	0.0000	0.9970
43	0.6645	0.3322	0.1839	0.8129	0.0000	0.9967	0.0000	0.9967	0.6645	0.3322	0.0000	0.9967
44	0.4982	0.4982	0.1655	0.8310	0.0000	0.9964	0.0000	0.9964	0.4982	0.4982	0.0000	0.9964
45	0.0000	0.9961	0.0000	0.9961	0.0000	0.9961	0.0000	0.9961	0.0000	0.9961	0.0000	0.9961

*AR(1) transition probabilities are from Krautmann, Ciecka, and Skoog (2010).

$$ {}^{aa}p_x^a > {}^{a}p_x^a > {}^{ia}p_x^a, \tag{1a} $$

$$ {}^{ia}p_x^a > {}^{i}p_x^a, \tag{1b} $$

$$ {}^{ai}p_x^a > {}^{i}p_x^a > {}^{ii}p_x^a \tag{1c} $$

These inequalities imply

$$ {}^{aa}p_x^i < {}^{a}p_x^i \tag{1a$'$} $$

$$ {}^{ia}p_x^i < {}^{i}p_x^i \tag{1b$'$} $$

$$ {}^{ai}p_x^i < {}^{i}p_x^i < {}^{ii}p_x^i \tag{1c$'$} $$

since death probabilities are equal in both the AR(1) and AR(2) models.[5] The observed relations among the AR(2) parameters, which we regard as most basic, are

$$ {}^{aa}p_x^a > {}^{ia}p_x^a > {}^{ai}p_x^a > {}^{ii}p_x^a. \tag{2} $$

Section IV shows that (2) implies (1a) and (1c). This leaves (1b), which also is proved in Section IV.

In addition, we observe the standard inequality for AR(1) models:

$$ {}^{a}p_x^a > {}^{i}p_x^a. \tag{2$'$} $$

Table 2 shows estimated worklife expectancies ${}^{aa}e_x^a$, ${}^{ia}e_x^a$, ${}^{ai}e_x^a$, and ${}^{ii}e_x^a$ for major league hitters based on our AR(2) model. The last two columns of this table contain corresponding AR(1) model expectancies ${}^{a}e_x^a$ and ${}^{i}e_x^a$. Typical relations[6] among the AR(2) population and estimated expectancies are given in inequality (3) and relations between the estimated AR(2) and estimated AR(1) expectancies in inequality (4).

$$ {}^{aa}e_x^a > {}^{ia}e_x^a > {}^{ai}e_x^a > {}^{ii}e_x^a \tag{3} $$

and

$$ {}^{i}e_x^a > {}^{ii}e_x^a \tag{4} $$

That is, being both active longer and more recently imply longer worklife [inequality (3)]; and shorter worklives occur for those who have been inactive longer [inequality (4)]. We also note that usually

$$ {}^{aa}e_x^a > {}^{a}e_x^a > {}^{ia}e_x^a. \tag{5} $$

We think of the middle term ${}^{a}e_x^a$ in inequality (5) as a blend of those who were active at both ages $x-1$ and x and those who were inactive at age $x-1$ but active at age x; thus ${}^{a}e_x^a$ falls between ${}^{aa}e_x^a$ and ${}^{ia}e_x^a$. We

Table 2. Worklife Expectancies for Major League Baseball Hitters with AR(2) and AR(1) Models*

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	${}^{aa}e_x^a$	${}^{ia}e_x^a$	${}^{ai}e_x^a$	${}^{ii}e_x^a$	${}^{a}e_x^a$	${}^{i}e_x^a$
23	8.32	8.01	5.32	3.31	8.13	6.30
24	7.74	7.05	3.83	3.31	7.45	5.38
25	7.21	6.21	2.99	2.03	6.88	4.37
26	6.64	5.43	2.29	1.51	6.36	3.50
27	6.31	5.05	1.74	1.17	6.01	2.87
28	5.97	4.07	1.36	0.96	5.64	2.35
29	5.56	3.49	0.96	0.85	5.25	2.02
30	5.25	3.65	1.02	0.55	4.95	1.65
31	4.85	3.17	0.75	0.38	4.58	1.24
32	4.35	2.99	0.65	0.21	4.18	0.89
33	3.92	2.91	0.49	0.07	3.85	0.59
34	3.63	2.82	0.20	0.04	3.58	0.40
35	3.29	2.38	0.12	0.00	3.24	0.27
36	2.93	2.11	0.22	0.00	2.90	0.23
37	2.74	1.95	0.08	0.00	2.71	0.15
38	2.51	1.79	0.12	0.00	2.53	0.12
39	2.37	1.68	0.07	0.00	2.36	0.09
40	2.15	1.58	0.10	0.00	2.18	0.07
41	1.98	1.41	0.11	0.00	2.16	0.04
42	1.80	1.41	0.00	0.00	1.99	0.00
43	2.00	1.28	0.00	0.00	2.00	0.00
44	1.50	1.17	0.00	0.00	1.50	0.00
45	1.00	1.00	0.00	0.00	1.00	0.00

*AR(1) worklife expectancies are from Krautmann, Ciecka, and Skoog (2010).

would have expected to have observed ${}^{ai}e_x^a > {}^{i}e_x^a > {}^{ii}e_x^a$, but inspection of Table 2 shows that this is not the case for the first part of the inequality ${}^{ai}e_x^a > {}^{i}e_x^a$, although it is true in our data for the second part ${}^{i}e_x^a > {}^{ii}e_x^a$.

Tables 3 and 4 contain distributional characteristics for major league hitters ages 23–45, and Figures 1–3 are pmfs for hitters at ages $x = 25$, 30, and 35 who were active at age x and either active or inactive at age $x - 1$. Tables 5 and 6 and Figures 4–6 are for hitters at the same ages $x = 25$, 30, and 35 who were inactive at age x and either active or inactive at age $x - 1$. These tables and graphs provide the first estimates

Table 3. Distributional Characteristics of Major League Baseball Hitters Based on AR(2) Model: Active for Year Prior to Listed Age and Active for Listed Age

(1) Age	(2) Mean	(3) Median	(4) Mode	(5) SD	(6) SK	(7) KU	(8) 10%	(9) 25%	(10) 75%	(11) 90%
23	8.32	8.00	5.00	4.51	0.43	2.45	3.00	5.00	12.00	15.00
24	7.74	7.00	4.00	4.35	0.47	2.47	2.00	4.00	11.00	14.00
25	7.21	7.00	3.00	4.17	0.50	2.49	2.00	4.00	10.00	13.00
26	6.64	6.00	2.00	4.02	0.51	2.51	2.00	3.00	10.00	12.00
27	6.31	6.00	3.00	3.79	0.52	2.56	2.00	3.00	9.00	12.00
28	5.97	6.00	2.00	3.56	0.53	2.64	2.00	3.00	9.00	11.00
29	5.56	5.00	2.00	3.34	0.57	2.73	1.00	3.00	8.00	10.00
30	5.25	5.00	4.00	3.09	0.62	2.88	1.00	3.00	7.00	10.00
31	4.85	4.00	3.00	2.85	0.71	3.04	1.00	3.00	7.00	9.00
32	4.35	4.00	2.00	2.67	0.79	3.19	1.00	2.00	6.00	8.00
33	3.92	4.00	1.00	2.49	0.85	3.35	1.00	2.00	5.00	7.00
34	3.63	3.00	1.00	2.28	0.94	3.58	1.00	2.00	5.00	7.00
35	3.29	3.00	2.00	2.09	1.05	3.82	1.00	2.00	4.00	6.00
36	2.93	2.00	1.00	1.94	1.12	3.96	1.00	1.00	4.00	6.00
37	2.74	2.00	1.00	1.77	1.16	4.07	1.00	1.00	4.00	5.00
38	2.51	2.00	1.00	1.63	1.19	4.14	1.00	1.00	3.00	5.00
39	2.37	2.00	1.00	1.47	1.23	4.20	1.00	1.00	3.00	4.00
40	2.15	2.00	1.00	1.33	1.29	4.14	1.00	1.00	3.00	4.00
41	1.98	2.00	1.00	1.21	1.26	3.60	1.00	1.00	2.00	4.00
42	1.80	1.00	1.00	1.10	1.00	2.44	1.00	1.00	3.00	4.00
43	2.00	2.00	1.00	0.82	0.01	1.50	1.00	1.00	3.00	3.00
44	1.50	1.00	1.00	0.50	0.01	1.00	1.00	1.00	2.00	2.00
45	1.00	1.00	1.00	0.00	——	——	1.00	1.00	1.00	1.00

of distributional characteristics from an AR(2) model. Distributions have positive skewness with means generally exceeding medians which in turn exceed modes. Figures 1–6 show generally declining probability mass values for additional years of activity, and Figures 4–6 show the dramatic impact of two consecutive years of inactivity on future careers.

Aggregate Measures for Major League Baseball Hitters

Formula (6) gives us the average remaining worklife (career) for active and inactive hitters, and formula (7) yields worklife for active

Table 4. Distributional Characteristics of Major League Baseball Hitters Based on AR(2) Model: Inactive for Year Prior to Listed Age and Active for Listed Age

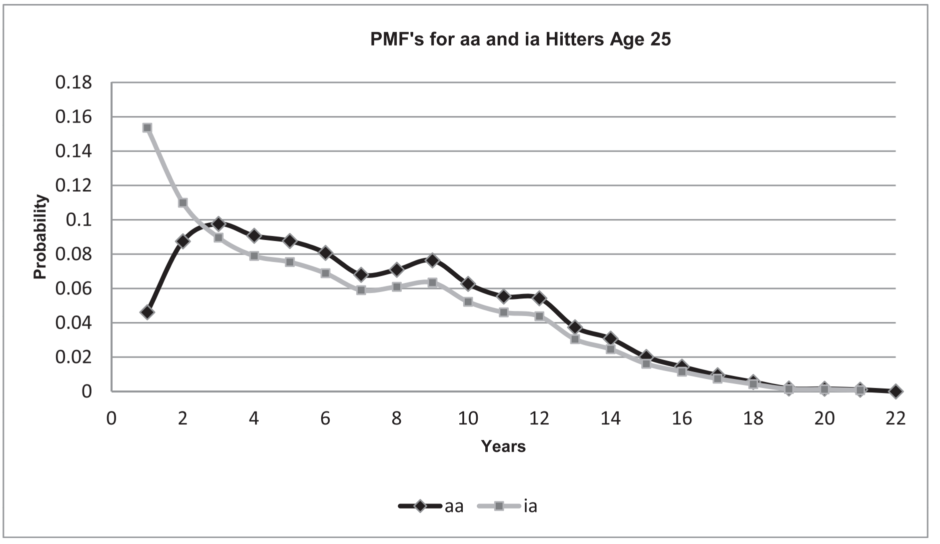
(1) Age	(2) Mean	(3) Median	(4) Mode	(5) SD	(6) SK	(7) KU	(8) 10%	(9) 25%	(10) 5%	(11) 90%
23	8.01	7.00	5.00	4.60	0.43	2.44	2.00	4.00	11.00	14.00
24	7.05	6.00	1.00	4.50	0.52	2.45	1.00	3.00	10.00	13.00
25	6.21	5.00	1.00	4.34	0.63	2.52	1.00	2.00	9.00	12.00
26	5.43	4.00	1.00	4.13	0.76	2.67	1.00	2.00	8.00	11.00
27	5.05	4.00	1.00	3.90	0.80	2.75	1.00	2.00	8.00	11.00
28	4.07	2.00	1.00	3.59	1.07	3.27	1.00	1.00	6.00	10.00
29	3.49	2.00	1.00	3.21	1.30	3.89	1.00	1.00	5.00	8.00
30	3.65	2.00	1.00	3.12	1.12	3.48	1.00	1.00	6.00	8.00
31	3.17	2.00	1.00	2.80	1.32	4.07	1.00	1.00	5.00	7.00
32	2.99	2.00	1.00	2.59	1.36	4.24	1.00	1.00	5.00	7.00
33	2.91	2.00	1.00	2.43	1.32	4.20	1.00	1.00	4.00	7.00
34	2.82	2.00	1.00	2.24	1.31	4.29	1.00	1.00	4.00	6.00
35	2.38	1.00	1.00	1.94	1.63	5.39	1.00	1.00	3.00	5.00
36	2.11	1.00	1.00	1.74	1.82	6.11	1.00	1.00	3.00	5.00
37	1.95	1.00	1.00	1.55	1.95	6.72	1.00	1.00	2.00	4.00
38	1.79	1.00	1.00	1.38	2.08	7.32	1.00	1.00	2.00	4.00
39	1.68	1.00	1.00	1.22	2.19	7.91	1.00	1.00	2.00	3.00
40	1.58	1.00	1.00	1.09	2.26	8.03	1.00	1.00	2.00	3.00
41	1.41	1.00	1.00	0.92	2.60	9.27	1.00	1.00	1.00	2.00
42	1.41	1.00	1.00	0.89	2.04	5.74	1.00	1.00	1.00	3.00
43	1.28	1.00	1.00	0.62	2.06	5.73	1.00	1.00	1.00	2.00
44	1.17	1.00	1.00	0.37	1.80	4.24	1.00	1.00	1.00	2.00
45	1.00	1.00	1.00	0.00	——	——	1.00	1.00	1.00	1.00

hitters:

$$\sum_x \left[\frac{{}^{aa}N_x\, {}^{aa}e_x^a + {}^{ia}N_x\, {}^{ia}e_x^a + {}^{ai}N_x\, {}^{ai}e_x^a + {}^{ii}N_x\, {}^{ii}e_x^a}{\sum_x ({}^{aa}N_x + {}^{ia}N_x + {}^{ai}N_x + {}^{ii}N_x)} \right] \qquad 23 \le x \le 45 \tag{6}$$

and

Figure 1.



$$\sum_x \left[\frac{{}^{aa}N_x \, {}^{aa}e_x^a + {}^{ia}N_x \, {}^{ia}e_x^a}{\sum_x ({}^{aa}N_x + {}^{ia}N_x)} \right] \qquad 23 \le x \le 45 \qquad (7)$$

where ${}^{aa}N_x \equiv {}^{aa}N_x^a + {}^{aa}N_x^i$, ${}^{ai}N_x \equiv {}^{ai}N_x^a + {}^{ai}N_x^i$, ${}^{ia}N_x \equiv {}^{ia}N_x^a + {}^{ia}N_x^i$ and

Figure 2.

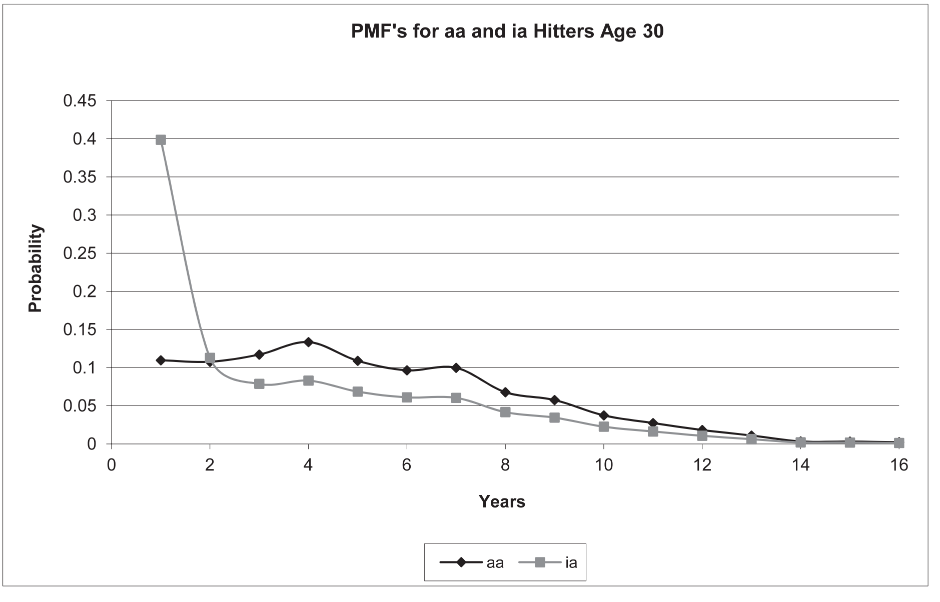
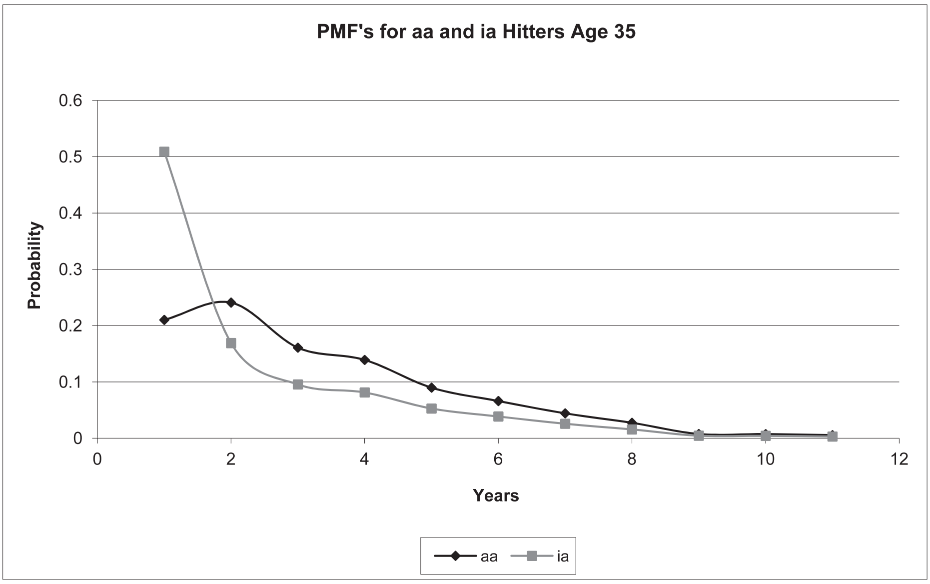


Figure 3.



${}^{ii}N_x \equiv {}^{ii}N_x^a + {}^{ii}N_x^i$ are counts of observations in various groups shown in the Appendix. Formula (6), the average remaining active time for hitters based on the AR(2) model and active and inactive players, yields 3.75 years; and the average based on only currently active players in formula (7) is 5.45 years.

The age and status-specific expectancies in Table 2 may be more useful in understanding career length than the high level of aggregation in formulae (6) and (7). For example, consider an active 30-year-old hitter who also was active at age $x - 1 = 29$ and who has a baseball worklife expectancy of ${}^{aa}e_{30}^{a} = 5.25$ years. In addition, we credit this player with one year of activity at age 29 and one year at age 30 – a total of 2 years for these ages. Suppose this player had accumulated another (say) 3 years of activity before age 29. Then, expected career length would be 10.25 years = (3 years prior to age 29 + 2 years for ages 29 and 30 + 5.25 years in expected future activity). A player similar in all respects (age, activity prior to age 29, activity status at age 29) but inactive at age 30 has an expected career of only 5.02 years = (3 years prior to age 29 + 1 year for age 29 + 1.02 years of expected future activity).

IV. Specification Error Analysis

Relations Among Transition Probabilities

The AR(2) model contains the AR(1) model as a special case, just as the AR(1) model contains the LP model (*i.e.*, the LPE model

Table 5. Distributional Characteristics of Major League Baseball Hitters Based on AR(2) Model: Active for Year Prior to Listed Age and Inactive for Listed Age

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age	**Mean**	**Median**	**Mode**	**SD**	**SK**	**KU**	**10%**	**25%**	**75%**	**90%**
23	5.32	4.00	0.00	4.72	0.72	2.61	0.00	1.00	9.00	12.00
24	3.83	2.00	0.00	4.31	1.10	3.32	0.00	0.00	7.00	11.00
25	2.99	1.00	0.00	3.86	1.39	4.15	0.00	0.00	5.00	9.00
26	2.29	1.00	0.00	3.41	1.70	5.24	0.00	0.00	3.00	8.00
27	1.74	0.00	0.00	2.91	2.07	6.90	0.00	0.00	2.00	6.00
28	1.36	0.00	0.00	2.53	2.38	8.56	0.00	0.00	1.00	5.00
29	0.96	0.00	0.00	2.15	2.86	11.49	0.00	0.00	1.00	4.00
30	1.02	0.00	0.00	2.13	2.71	10.59	0.00	0.00	1.00	4.00
31	0.75	0.00	0.00	1.81	3.16	13.77	0.00	0.00	1.00	3.00
32	0.65	0.00	0.00	1.65	3.34	15.16	0.00	0.00	0.00	2.00
33	0.49	0.00	0.00	1.41	3.75	18.74	0.00	0.00	0.00	1.00
34	0.20	0.00	0.00	0.86	6.07	46.67	0.00	0.00	0.00	0.00
35	0.12	0.00	0.00	0.64	7.69	74.58	0.00	0.00	0.00	0.00
36	0.22	0.00	0.00	0.80	5.38	38.23	0.00	0.00	0.00	1.00
37	0.08	0.00	0.00	0.49	8.40	90.78	0.00	0.00	0.00	0.00
38	0.12	0.00	0.00	0.53	6.75	60.34	0.00	0.00	0.00	0.00
39	0.07	0.00	0.00	0.40	8.14	86.73	0.00	0.00	0.00	0.00
40	0.10	0.00	0.00	0.44	6.43	55.91	0.00	0.00	0.00	0.00
41	0.11	0.00	0.00	0.45	5.68	41.36	0.00	0.00	0.00	0.00
42	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
43	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
44	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
45	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00

without considering employment probabilities). Thus, one may test the AR(1) model, informally or formally (by testing whether individual or groups of coefficients are zero).[6] One may also, with the aid of algebra and laws of large numbers, interpret parameters of the AR(1) model fit to data in light of the AR(2) parameters. An example of this latter procedure, the connection between the standard AR(2) and AR(1) time series model parameters, appears in Kiefer and Skoog (1984).

The four essential transition probabilities of the AR(2) model are ${}^{aa}p_x^a$, ${}^{ai}p_x^a$, ${}^{ia}p_x^a$ and ${}^{ii}p_x^a$. As these terms appear, they have been reduced for mortality probability (as defined in GC4), which is

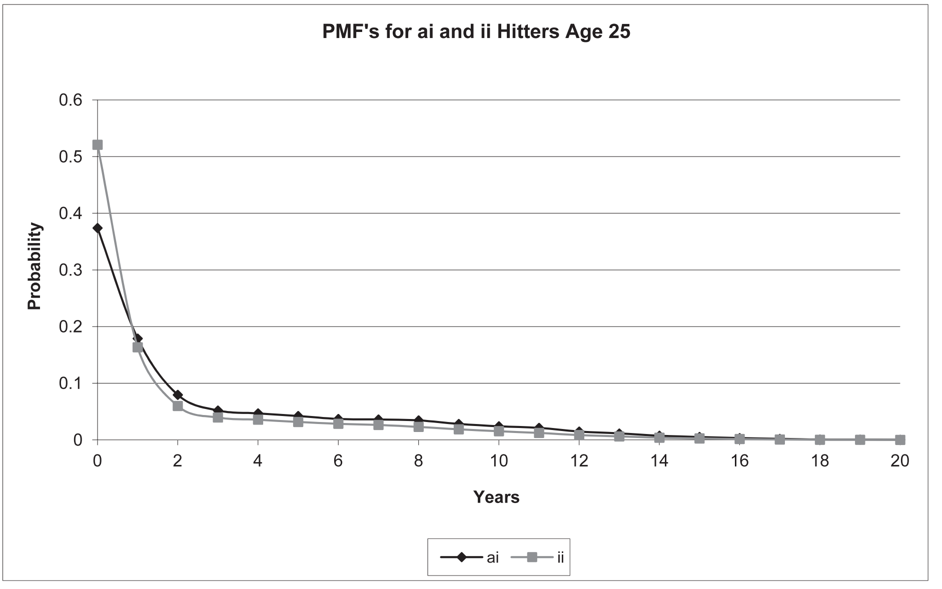
Table 6. Distributional Characteristics of Major League Baseball Hitters Based on AR(2) Model: Inactive for Year Prior to Listed Age and Inactive for Listed Age

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age	**Mean**	**Median**	**Mode**	**SD**	**SK**	**KU**	**10%**	**25%**	**75%**	**90%**
23	3.31	1.00	0.00	4.12	1.29	3.83	0.00	0.00	6.00	10.00
24	3.31	1.00	0.00	4.12	1.29	3.82	0.00	0.00	6.00	10.00
25	2.03	0.00	0.00	3.29	1.92	6.18	0.00	0.00	3.00	7.00
26	1.51	0.00	0.00	2.80	2.30	8.08	0.00	0.00	1.00	6.00
27	1.17	0.00	0.00	2.41	2.63	10.06	0.00	0.00	1.00	5.00
28	0.96	0.00	0.00	2.16	2.89	11.77	0.00	0.00	1.00	4.00
29	0.85	0.00	0.00	2.02	3.09	13.11	0.00	0.00	1.00	3.00
30	0.55	0.00	0.00	1.60	3.87	19.65	0.00	0.00	0.00	1.00
31	0.38	0.00	0.00	1.32	4.62	27.23	0.00	0.00	0.00	1.00
32	0.21	0.00	0.00	0.96	6.29	48.85	0.00	0.00	0.00	0.00
33	0.07	0.00	0.00	0.53	10.66	138.35	0.00	0.00	0.00	0.00
34	0.04	0.00	0.00	0.42	13.08	208.66	0.00	0.00	0.00	0.00
35	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
36	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
37	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
38	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
39	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
41	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
42	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
43	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
44	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00
45	0.00	0.00	0.00	0.00	——	——	0.00	0.00	0.00	0.00

assumed independent of the initial states (as defined in GC3): ${}^{aa}p_x^d = {}^{ai}p_x^d = {}^{ia}p_x^d = {}^{ii}p_x^d \equiv {}^{\bullet}p_x^d$. We estimate ${}^{\bullet}p_x^d$ from an outside source, since the sample of active players dying over the age groups in question is exceedingly small. Finally, the transitions into inactivity ${}^{aa}p_x^i$, ${}^{ai}p_x^i$, ${}^{ia}p_x^i$ and ${}^{ii}p_x^i$ are determined from the essential transition probabilities and ${}^{\bullet}p_x^d$ by GC4.

Inequality (2) records that, within the AR(2) model, ${}^{aa}p_x^a > {}^{ia}p_x^a > {}^{ai}p_x^a > {}^{ii}p_x^a$; the first term dominates all the other terms saying that more past activity results in more future activity while the relation between the middle terms says more recent activity dominates

Figure 4.



earlier activity in regard to future activity. Even though the data favor the AR(2) model, we can still mechanically fit an AR(1) model to the data. In Section III we recorded some of the empirical findings when this is done. In this section, we elaborate and attempt to interpret these results. The AR(1) model says simply that the AR(2) model obeys two linear restrictions, ${}^{aa}p_x^a = {}^{ia}p_x^a$ and that ${}^{ai}p_x^a = {}^{ii}p_x^a$. If these are met, we let

Figure 5.

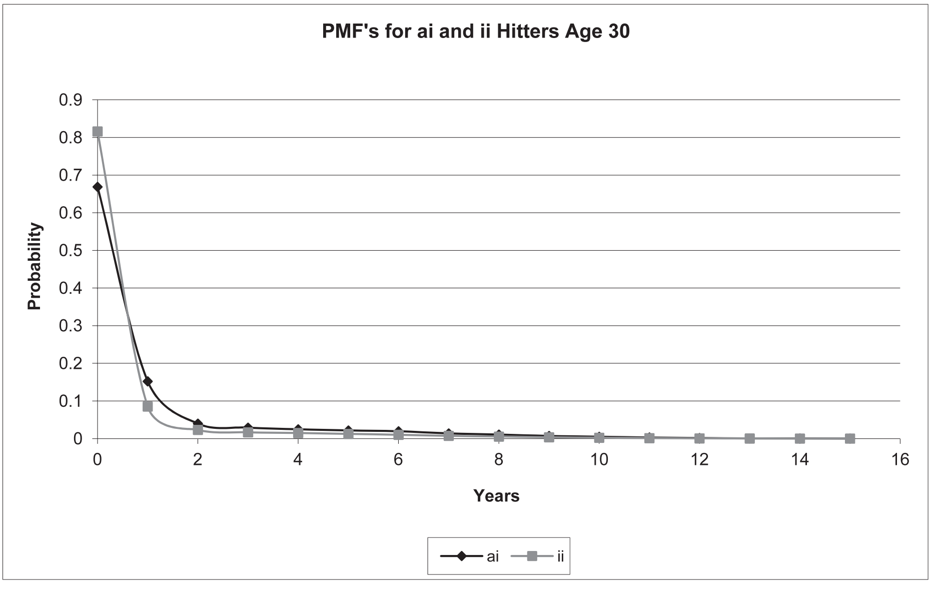
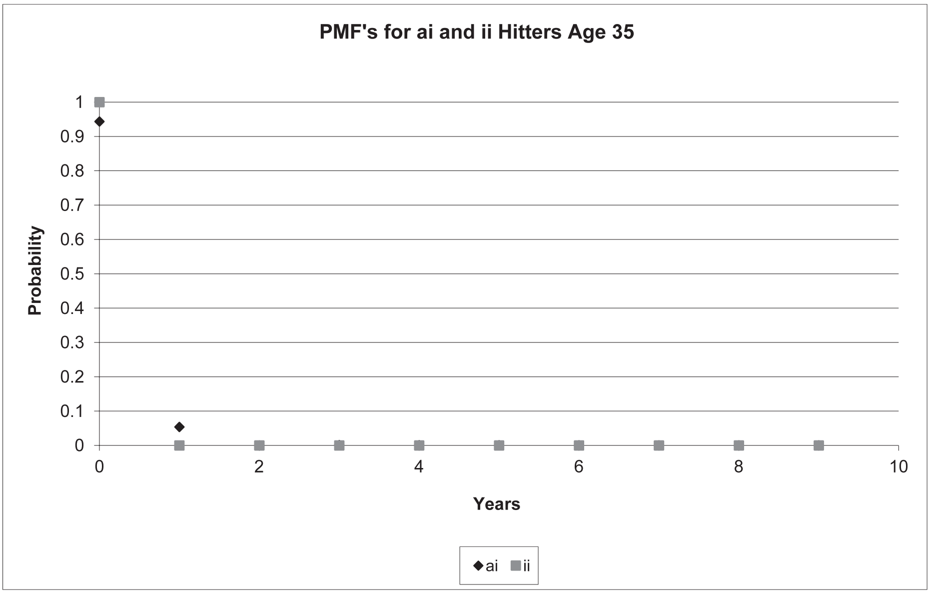


Figure 6.



${}^{aa}p_x^a = {}^{ia}p_x^a \equiv {}^{a}p_x^a$ and ${}^{ai}p_x^a = {}^{ii}p_x^a \equiv {}^{i}p_x^a$; the result is an AR(1) model with these parameters. Glancing down Columns 2 and 4 of Table 1, it is apparent, as noted in our first inequality in (2), that the restrictions implied by the AR(1) model would not pass a formal statistical hypothesis test, at least at ages 25 – 35. If a player is active two years in a row, he has a much higher chance of being active next year than if he was inactive two years ago but is active this year. Similarly, given that a player is currently inactive, his chances of appearing in a major league game next year are better if he was active last year (see Table 1, columns 6 and 8). We say that there is additional information in yesterday's state beyond that contained in today's state.

In this section, it will be useful to re-interpret the four essential transition probabilities ${}^{aa}p_x^a$, ${}^{ai}p_x^a$, ${}^{ia}p_x^a$ and ${}^{ii}p_x^a$ both as being conditional on survival and as referring to their sampling counterparts. We do this for ease of notation.[7] Our basic data consist of counts. Our notation is that ${}^{km}N_x^n$ represents the number of players in state k at $x-1$ and in state m at x who transition to state n at age x+1. Our AR(2) estimates are then

$$ {}^{aa}p_x^a = \frac{{}^{aa}N_x^a}{{}^{aa}N_x^a + {}^{aa}N_x^i}, \qquad {}^{ia}p_x^a = \frac{{}^{ia}N_x^a}{{}^{ia}N_x^a + {}^{ia}N_x^i}, $$

$$ {}^{ai}p_x^a = \frac{{}^{ai}N_x^a}{{}^{ai}N_x^a + {}^{ai}N_x^i}, \quad \text{and} \quad {}^{ii}p_x^a = \frac{{}^{ii}N_x^a}{{}^{ii}N_x^a + {}^{ii}N_x^i}. \tag{8} $$

When we fit an AR(1) model, we ignore the state information for period $x-1$, so that we define the (collapsed) counts usually used in the increment-decrement model as ${}^{a}N_x^a \equiv {}^{aa}N_x^a + {}^{ia}N_x^a$, ${}^{i}N_x^a \equiv {}^{ai}N_x^a + {}^{ii}N_x^a$, ${}^{a}N_x^i \equiv {}^{aa}N_x^i + {}^{ia}N_x^i$ and ${}^{i}N_x^i \equiv {}^{ai}N_x^i + {}^{ii}N_x^i$. The usual AR(1) estimators are then:

$$
{}^{a}p_x^a = \frac{{}^{a}N_x^a}{{}^{a}N_x^a + {}^{a}N_x^i}, \qquad {}^{i}p_x^a = \frac{{}^{i}N_x^a}{{}^{i}N_x^a + {}^{i}N_x^i},
$$

$$
{}^{a}p_x^i = \frac{{}^{a}N_x^i}{{}^{a}N_x^i + {}^{a}N_x^a}, \quad \text{and} \quad {}^{i}p_x^i = \frac{{}^{i}N_x^i}{{}^{i}N_x^i + {}^{i}N_x^a}, \tag{9}
$$

with more compact notation for denominators being ${}^{a}N_x = {}^{a}N_x^a + {}^{a}N_x^i$ and ${}^{i}N_x = {}^{i}N_x^a + {}^{i}N_x^i$. Now, we may write:

$$
{}^{a}p_x^a = \frac{{}^{a}N_x^a}{{}^{a}N_x^a + {}^{a}N_x^i} = \frac{{}^{aa}N_x^a + {}^{ia}N_x^a}{{}^{a}N_x} = \frac{{}^{aa}N_x^a}{{}^{a}N_x} + \frac{{}^{ia}N_x^a}{{}^{a}N_x}
$$

$$
= \frac{{}^{aa}N_x^a}{{}^{aa}N_x}\frac{{}^{aa}N_x}{{}^{a}N_x} + \frac{{}^{ia}N_x^a}{{}^{ia}N_x}\frac{{}^{ia}N_x}{{}^{a}N_x} \tag{10a}
$$

$$
= {}^{aa}p_x^a \frac{{}^{aa}N_x}{{}^{a}N_x} + {}^{ia}p_x^a \frac{{}^{ia}N_x}{{}^{a}N_x} \equiv {}^{aa}p_x^a w_{1,x} + {}^{ia}p_x^a (1 - w_{1,x}) \tag{10b}
$$

and, in a similar manner,

$$
{}^{a}p_x^i = {}^{aa}p_x^i w_{1,x} + {}^{ia}p_x^i (1 - w_{1,x}) \tag{11}
$$

where $w_{1,x} \equiv {}^{aa}N_x / {}^{a}N_x$ is the fraction of those active at x who also were active at $x - 1$. Consequently, since $0 < w_{1,x} < 1$ generally, the AR(1) estimated transition probability ${}^{a}p_x^a$ will be between the AR(2) probabilities ${}^{aa}p_x^a$ and ${}^{ia}p_x^a$. Given the empirical finding that ${}^{aa}p_x^a > {}^{ia}p_x^a$, it follows that ${}^{aa}p_x^a > {}^{a}p_x^a > {}^{ia}p_x^a$, establishing (1a).

Also, in the AR(1) model, we have

$$
{}^{i}p_x^a = \frac{{}^{i}N_x^a}{{}^{i}N_x^a + {}^{i}N_x^i} = \frac{{}^{ai}N_x^a + {}^{ii}N_x^a}{{}^{i}N_x} = \frac{{}^{ai}N_x^a}{{}^{i}N_x} + \frac{{}^{ii}N_x^a}{{}^{i}N_x} = \frac{{}^{ai}N_x^a}{{}^{ai}N_x}\frac{{}^{ai}N_x}{{}^{i}N_x} + \frac{{}^{ii}N_x^a}{{}^{ii}N_x}\frac{{}^{ii}N_x}{{}^{i}N_x} \tag{12a}
$$

$$
= {}^{ai}p_x^a \frac{{}^{ai}N_x}{{}^{i}N_x} + {}^{ii}p_x^a \frac{{}^{ii}N_x}{{}^{i}N_x} \equiv {}^{ai}p_x^a w'_{1,x} + {}^{ii}p_x^a (1 - w'_{1,x}) \tag{12b}
$$

and, in a similar manner,

$$
{}^{i}p^i = {}^{ai}p_x^i w'_{1,x} + {}^{ii}p_x^i (1 - w'_{1,x}) \tag{13}
$$

where $w'_{1,x} \equiv {}^{ai}N_x / {}^{i}N_x$ is the fraction of those inactive at x who also were active at $x - 1$. Since again $0 < w'_{1,x} < 1$ generally, the AR(1) estimated transition probability ${}^{i}p_x^a$ will be will be between the AR(2) probabilities ${}^{ai}p_x^a$ and ${}^{ii}p_x^a$. Given the empirical finding that ${}^{ai}p_x^a > {}^{ii}p_x^a$, it follows that ${}^{ai}p_x^a > {}^{i}p_x^a > {}^{ii}p_x^a$, establishing (1c).

Finally, since from (2) ${}^{ia}p_x^a > {}^{ai}p_x^a > {}^{ii}p_x^a$ (*i.e.*, ${}^{ia}p_x^a$ dominates each of the last two terms whose weighted average has just been shown in (12) to be ${}^{i}p_x^a$), then (1b), ${}^{ia}p_x^a > {}^{i}p_x^a$, is proved. We note that proving (2′) would require the middle inequality in the expression
${}^{a}p_x^a = {}^{aa}p_x^a w_{1,x} + {}^{ia}p_x^a(1 - w_{1,x}) > {}^{ai}p_x^a w_{1,x}^{'} + {}^{ii}p_x^a(1 - w'_{1,x}) = {}^{i}p_x^a$.
If $w_{1,x} = w_{1,x}^{'}$ the result would follow, but since do not expect that the likelihood of previous activity to be the same whether currently active or not, the proof breaks down at this point.

Interpretation of AR(1) Transition Probabilities

Let us re-define the previous weights with the suggestive notation

$$\tilde{p}(a_{x-1}|a_x) \equiv w_{1,x} \equiv \frac{{}^{aa}N_x}{{}^{a}N_x}, \qquad \tilde{p}(i_{x-1}|a_x) \equiv 1 - w_{1,x} \equiv \frac{{}^{ia}N_x}{{}^{a}N_x} \tag{14a}$$

$$\tilde{p}(a_{x-1}|i_x) \equiv w'_{1,x} \equiv \frac{{}^{ai}N_x}{{}^{i}N_x}, \qquad \tilde{p}(i_{x-1}|i_x) \equiv 1 - w'_{1,x} \equiv \frac{{}^{ii}N_x}{{}^{i}N_x}. \tag{14b}$$

For example, the left hand side terms are the "backcasts" estimates of the probability that a person observed active at x, a_x, will have been active at $x - 1$, a_{x-1}. As sample sizes ${}^{aa}N_x$, ${}^{ai}N_x$, ${}^{ia}N_x$ and ${}^{ii}N_x$ grow, $\tilde{p}(a_{x-1}|a_x) \rightarrow p(a_{x-1}|a_x)$ in whatever probability sense one wishes, where of course $p(a_{x-1}|a_x)$ refers to the population probability that a person observed active at x, a_x, will have been active at $x - 1$. The AR(1) transition probabilities thus have the following common sense interpretation. Assume that a person is observed at age x to be active, and you wish to estimate the probability that he will be active at age x+1. If we knew that he had been active at $x - 1$, ${}^{aa}p_x^a$ is the answer. If we knew that he had been inactive at $x - 1$, ${}^{ia}p_x^a$ is the answer. Since we do not know which is correct, we estimate the probability of each alternative and use the laws of conditional probability using (10b) and the interpretation of $w_{1,x}$ in (14a). These hold in the sample and in the limit (*i.e.*, in the population).

The interpretation of AR(1) in an AR(2) world is that it is optimizing; it provides the best – in the sense of unbiased and least variance – probability predictions, given the limited information available to it. This is the same message, here for discrete valued discrete time AR(2) and AR(1) stochastic processes, that was demonstrated by Kiefer and Skoog (1984) for the more usual continuous valued discrete time AR(2) and AR(1) time series $x_{t+1} = \phi_1 x_t + \phi_2 x_{t-1} + \mu_t$ that are studied in econometrics.

Relations Among Expectations

Since worklife expectancies are very complicated functions of the underlying transition probabilities, the empirical inequalities of (1), (1′), (2), and (2′) might be expected to carry over to the analogous expectancies. The remainder of this section proves some results and gives an interpretation for why other results we expect cannot be proven.

Skoog (2003) showed that the four AR(2) worklife expectancies at age x appearing on the left hand sides of the equations immediately below are determined recursively by the right hand sides (once again assuming mid-period transitions):

AR(2)-1 $\quad {}^{aa}e_x^a = .5 + .5\,{}^{aa}p_x^a + {}^{aa}p_x^a\,{}^{aa}e_{x+1}^a + {}^{aa}p_x^i\,{}^{ai}e_{x+1}^a$

AR(2)-2 $\quad {}^{ia}e_x^a = .5 + .5\,{}^{ia}p_x^a + {}^{ia}p_x^a\,{}^{aa}e_{x+1}^a + {}^{ia}p_x^i\,{}^{ai}e_{x+1}^a$

AR(2)-3 $\quad {}^{ai}e_x^a = .5\,{}^{ai}p_x^a + {}^{ai}p_x^a\,{}^{ia}e_{x+1}^a + {}^{ai}p_x^i\,{}^{ii}e_{x+1}^a$

AR(2)-4 $\quad {}^{ii}e_x^a = .5\,{}^{ii}p_x^a + {}^{ii}p_x^a\,{}^{ia}e_{x+1}^a + {}^{ii}p_x^i\,{}^{ii}e_{x+1}^a.$

Recalling inequality (2) above, ${}^{aa}p_x^a > {}^{ia}p_x^a > {}^{ai}p_x^a > {}^{ii}p_x^a$, the first inequality ${}^{aa}p_x^a > {}^{ia}p_x^a$ implies that ${}^{aa}e_x^a > {}^{ia}e_x^a$ for $x < TA - 1$. We can see this result in the following manner. At age $TA - 1$, ${}^{aa}e_{TA-1}^a = {}^{ia}e_{TA-1}^a = .5$; but we wish to show that ${}^{aa}e_x^a > {}^{ia}e_x^a$ for $x < TA - 1$. At ages $x < TA - 1$, each term on the right hand side of AR(2)-1 exceeds its counterpart in AR(2)-2 (except for the .5 term which is common to both recursions). Therefore, ${}^{aa}e_x^a > {}^{ia}e_x^a$ (*i.e.*, the first part of inequality (3)) when ${}^{aa}p_x^a > {}^{ia}p_x^a$ for $x < TA - 1$. Similarly, term-by-term comparisons of the right-hand-sides of AR(2)-3 and AR(2)-4 establish ${}^{ai}e_x^a > {}^{ii}e_x^a$(the very last part of inequality (3)) when ${}^{ai}p_x^a > {}^{ii}p_x^a$ for $x < TA - 1$ and ${}^{ai}e_{TA-1}^a = {}^{ii}e_{TA-1}^a = 0$. Finally, although we expect to usually observe ${}^{ia}e_x^a > {}^{ai}e_x^a$ (*i.e.*, the middle of inequality (3)) when inequality (2) holds, it is not a mathematical certainty. The interaction between transition probabilities and worklife expectancies can lead to ${}^{ia}e_x^a < {}^{ai}e_x^a$ in some cases.

Skoog (2002) showed that for an AR(1) model, the two AR(1) worklife expectancies at age x appearing on the left hand sides of the equations immediately below are determined recursively by the right hand sides:

AR(1)-1 $\quad {}^{a}e_x^a = .5 + .5\,{}^{a}p_x^a + {}^{a}p_x^a\,{}^{a}e_{x+1}^a + {}^{a}p_x^i\,{}^{i}e_{x+1}^a$

AR(1)-2 $\quad {}^{i}e_x^a = .5\,{}^{i}p_x^a + {}^{i}p_x^a\,{}^{a}e_{x+1}^a + {}^{i}p_x^i\,{}^{i}e_{x+1}^a.$

Term-by-term comparisons between AR(1)-1 and AR(1)-2 establish

${}^{a}e_x^a > {}^{i}e_x^a$ in the same manner in which we proved ${}^{aa}e_x^a > {}^{ia}e_x^a$. These equations hold both in the population and in any sample when the true model is AR(1). When the true model is not AR(1), but AR(2) or something else, we may contemplate fitting a mis-specified AR(1) to the data. In this case, we should introduce analogous symbols to reflect both the fact that the equation contains estimates and that there is specification error. Thus, we have:

AR(1)-1 estimated, mis-specified $\quad {}^{a}\tilde{e}_x^a = .5 + .5\,{}^{a}\tilde{p}_x^a + {}^{a}\tilde{p}_x^a\,{}^{a}\tilde{e}_{x+1}^a + {}^{a}\tilde{p}_x^i\,{}^{i}\tilde{e}_{x+1}^a$

AR(1)-2 estimated, mis-specified $\quad {}^{i}\tilde{e}_x^a = .5\,{}^{i}\tilde{p}_x^a + {}^{i}\tilde{p}_x^a\,{}^{a}\tilde{e}_{x+1}^a + {}^{i}\tilde{p}_x^i\,{}^{i}\tilde{e}_{x+1}^a.$

Columns (6) and (7) of Table 2 (written without the $\sim$) report such ${}^{a}\tilde{e}_x^a$ and ${}^{i}\tilde{e}_x^a$; the force of these equations is that the same recursions as in AR(1)-1 and AR(1)-2 hold under mis-specification.

Multiplying the AR(2)-1 equation by $w_{1,x}$ and the AR(2)-2 equation by $1 - w_{1,x}$ results in the population equation, if desired, estimating equations

$$w_{1,x}\,{}^{aa}e_x^a = .5w_{1,x} + .5w_{1,x}\,{}^{aa}p_x^a + w_{1,x}\,{}^{aa}p_x^a\,{}^{aa}e_{x+1}^a + w_{1,x}\,{}^{aa}p_x^i\,{}^{ai}e_{x+1}^a \quad \text{and} \tag{15a}$$

$$\begin{aligned}(1 - w_{1,x})^{ia}e_x^a = {} & .5(1 - w_{1,x}) + .5(1 - w_{1,x})^{ia}p_x^a \\ & + (1 - w_{1,x})^{ia}p_x^a\,{}^{aa}e_{x+1}^a + (1 - w_{1,x})^{ia}p_x^i\,{}^{ai}e_{x+1}^a.\end{aligned} \tag{15b}$$

Adding these, defining ${}^{a}\bar{e}_x^a \equiv w_{1,x}\,{}^{aa}e_x^a + (1 - w_{1,x})^{ia}e_x^a$ and recalling that ${}^{a}p_x^a \equiv {}^{aa}p_x^a w_{1,x} + {}^{ia}p_x^a(1 - w_{1,x})$ and ${}^{a}p_x^i = {}^{aa}p_x^i w_{1,x} + {}^{ia}p_x^i(1 - w_{1,x})$ from (10b) and (11), results in the AR(1) pseudo-recursion:[8]

$${}^{a}\bar{e}_x^a = .5 + .5\,{}^{a}p_x^a + {}^{a}p_x^a\,{}^{aa}e_{x+1}^a + {}^{a}p_x^i\,{}^{ai}e_{x+1}^a. \tag{15c}$$

Now, we multiply AR(2)-3 and AR(2)-4 recursions by $w_{1,x}$ and $1 - w'_{1,x}$, respectively, to obtain

$$w'_{1,x}\,{}^{ai}e_x^a = .5w'_{1,x}\,{}^{ai}p_x^a + w'_{1,x}\,{}^{ai}p_x^a\,{}^{ia}e_{x+1}^a + w'_{1,x}\,{}^{ai}p_x^i\,{}^{ii}e_{x+1}^a \quad \text{and} \tag{16a}$$

$$\begin{aligned}(1 - w'_{1,x})^{ii}e_x^a = {} & .5(1 - w'_{1,x})^{ii}p_x^a + (1 - w'_{1,x})^{ii}p_x^a\,{}^{ia}e_{x+1}^a \\ & + (1 - w'_{1,x})^{ii}p_x^i\,{}^{ii}e_{x+1}^a.\end{aligned} \tag{16b}$$

Recalling ${}^{i}p_x^a = {}^{ai}p_x^a w'_{1,x} + {}^{ii}p_x^a(1 - w'_{1,x})$ and ${}^{i}p_x^i = {}^{ai}p_x^i w'_{1,x} + {}^{ii}p_x^i(1 - w'_{1,x})$ from (12b) and (13), a second AR(1) pseudo-recursion becomes:

$${}^{i}\bar{e}_x^a = .5\,{}^{i}p_x^a + {}^{i}p_x^a\,{}^{ia}e_{x+1}^a + {}^{i}p_x^i\,{}^{ii}e_{x+1}^a. \tag{16c}$$

The weighted averages ${}^{a}\bar{e}_x^a \equiv w_{1,x}\,{}^{aa}e_x^a + (1 - w_{1,x})^{ia}e_x^a$ and ${}^{i}\bar{e}_x^a \equiv w'_{1,x}\,{}^{ai}e_x^a + (1 - w'_{1,x})^{ii}e_x^a$ equal (15c) and (16c), respectively, which use AR(1) transition probabilities. Since $0 \leq w_{1,x} \leq 1$ and $0 \leq w'_{1,x} \leq 1$, both the strengthened version of inequality (4) ${}^{ii}e^a_x < {}^{i}\bar{e}^a_x < {}^{ai}e^a_x$ and the version of inequality (5) ${}^{aa}e^a_x > {}^{a}\bar{e}^a_x > {}^{ia}e^a_x$ hold. It is apparent, however, by comparing the right hand sides of the recursions involving ${}^{a}\tilde{e}^a_x$ and ${}^{i}\tilde{e}^a_x$ with (15c) and (16c), that ${}^{a}\bar{e}^a_x \neq {}^{a}\tilde{e}^a_x$ and ${}^{i}\bar{e}^a_x \neq {}^{i}\tilde{e}^a_x$, so that the pseudorecursion estimated values are different from the AR(1) mis-specified values. In other words, while the $w_{1,x}$ and $w'_{1,x}$ weights are correct for forming the AR(1) transition probabilities, they are inappropriate for capturing the ${}^{a}e^a_x$ and ${}^{i}e^a_x$ estimated values.

We turn to an explanation as to why we cannot obtain the hoped for estimation inequalities (ignoring use of the $\sim$) ${}^{ai}e^a > {}^{i}e^a_x > {}^{ii}e^a_x$, a stronger form of (4), and ${}^{aa}e^a_x > {}^{a}e^a_x > {}^{ia}e^a_x$ in (5) connecting the AR(2) and AR1 expectancies. Even if it were true that ${}^{aa}e^a_{x+1} > {}^{a}e^a_{x+1}$ and ${}^{ai}e^a_{x+1} > {}^{i}e^a_{x+1}$, the attempt to conclude that ${}^{aa}e^a_x = .5 + .5\,{}^{aa}p^a_x + {}^{aa}p^a_x\,{}^{aa}e^a_{x+1} + {}^{aa}p^i_x\,{}^{ai}e^a_{x+1}$ exceeds ${}^{a}e^a_x = .5 + .5\,{}^{a}p^a_x + {}^{a}p^a_x\,{}^{aa}e^a_{x+1} + {}^{a}p^i_x\,{}^{i}e^a_{x+1}$ fails because it is *not* true that ${}^{aa}p^i_x > {}^{a}p^i_x$, since from (1′) ${}^{aa}p^i_x < {}^{a}p^i_x$. The other AR(2) and AR(1) comparisons fail for the same reason – the transition probability into the *i* state precludes an inequality like (5) from always holding, although empirically the result often holds.

We note the simplifications of the AR(2) recursions if the model is truly AR(1). In this case, let ${}^{aa}p^a_x = {}^{ia}p^a_x \equiv {}^{a}p^a_x$ and ${}^{ai}p^a_x = {}^{ii}p^a_x \equiv {}^{i}p^a_x$. Substituting into the AR(2) recursions yields the following specialized recursions:

AR(2)-1 specialized $\quad {}^{aa}e^a_x = .5 + .5\,{}^{a}p^a_x + {}^{a}p^a_x\,{}^{aa}e^a_{x+1} + {}^{a}p^i_x\,{}^{ai}e^a_{x+1}$

AR(2)-2 specialized $\quad {}^{ia}e^a_x = .5 + .5\,{}^{a}p^a_x + {}^{a}p^a_x\,{}^{aa}e^a_{x+1} + {}^{a}p^i_x\,{}^{ai}e^a_{x+1}$

AR(2)-3 specialized $\quad {}^{ai}e^a_x = .5\,{}^{i}p^a_x + {}^{i}p^a_x\,{}^{ia}e^a_{x+1} + {}^{i}p^i_x\,{}^{ii}e^a_{x+1}$

AR(2)-4 specialized $\quad {}^{ii}e^a_x = .5\,{}^{i}p^a_x + {}^{i}p^a_x\,{}^{ia}e^a_{x+1} + {}^{i}p^i_x\,{}^{ii}e^a_{x+1}$

Note that AR(2)-1 specialized and AR(2)-2 specialized have the same right hand sides, so that ${}^{aa}e^a_x = {}^{ia}e^a_x \equiv {}^{a}e^a_x$; similarly, AR(2)-3 specialized and AR(2)-4 specialized have the same right hand sides, so that ${}^{ai}e^a_x = {}^{ii}e^a_x \equiv {}^{i}e^a_x$. Eliminating the redundant AR(2)-2 and AR(2)-4 results in the familiar AR(1) recursions.

V. Conclusion

Recursions defining entire probability mass functions for an autoregressive model of order two have been specified – a first step in moving beyond the one-period memory of a Markov process. These recursions await national representative data on age, gender, labor force status, and educational attainment in order to become operational for most forensic purposes. Worklife expectancy recursions for the AR(2) and fitted AR(1) models were provided, along with stylized inequalities involving the AR(2) transition probabilities. Most transition inequalities within the AR(2) and AR(1) models carried over to their worklife expectancies, but transition inequalities across the AR(2) and AR(1) models did not carry over generally to the worklife expectancies. We applied the AR(2) probability mass and worklife expectancy recursions to major league baseball hitters and estimated worklife expectancies and other distributional characteristics. Expectancies with an AR(1) model were shown as well; empirical relations between AR(2) and AR(1) are summarized in inequalities (1)-(5). In particular, when we exclude ages beyond 37 because data are very thin, ${}^{aa}e_x^a > {}^{a}e_x^a > {}^{ia}e_x^a$ and ${}^{ii}e_x^a < {}^{i}e_x^a$. We note that ${}^{ii}e_x^a$ and ${}^{i}e_x^a$ differ more than ${}^{aa}e_x^a$ and ${}^{a}e_x^a$. For example, at age 25, ${}^{ii}e_{25}^a = 2.03$ years and ${}^{i}e_{25}^a = 4.37$ years – a difference of 2.34 years; but ${}^{aa}e_{25}^a = 7.21$ years and ${}^{a}e_{25}^a = 6.88$ years differ by only .33 years. Of course, the relative closeness of ${}^{aa}e_x^a$ and ${}^{a}e_x^a$ and the larger differences between ${}^{ii}e_x^a$ and ${}^{i}e_x^a$ may be peculiar to baseball; but these results could signal a feature of labor force data in general.

References

Bureau of Labor Statistics. 1982. *Tables of Working Life: The Increment-Decrement Model, Bulletin 2135*.

________. 1986. *Worklife Estimates: Effects of Race and Education, Bulletin 2254*.

Cushing, Matthew J. and David I. Rosenbaum. 2011. Higher Order Markov Estimates of Worklife: Comparing SIPP to CPS Results, paper presented at the Western Economics Association meeting, San Diego, CA.

Gillette, Gary and Pete Palmer. 2008. *The ESPN Baseball Encyclopedia*, Fifth Edition, Sterling Publishing Company, New York: NY.

James, Bill, *Major League Handbook, 1998 – 2008*, STATS Publishing, Skokie: IL.

Kiefer, Nicholas and Gary R. Skoog. 1984. Local Asymptotic Specification Error Analysis. *Econometrica*, 52(4), 873–885.

Krautmann, Anthony C., James E. Ciecka, and Gary R. Skoog. 2010. A Markov Process Model of the Number of Years Spent in Major League Baseball. *Journal of Quantitative Analysis in Sports*, 6(4), Article 6, 1–23.

Skoog, Gary R. 2002. Worklife Expectancy: Theoretical Results. paper presented at the ASSA meeting, Atlanta, Georgia.

Skoog, Gary R. 2003. Generalization of the Increment-Decrement Model. paper presented at the ASSA meeting, Washington, D.C.

Skoog, Gary R. and James E. Ciecka. 2001a. The Markov (Increment-Decrement) Model of Labor Market Activity: New Results Beyond Worklife Expectancies. *Journal of Legal Economics*, 11(1), 1–21.

________. 2001b. The Markov (Increment-Decrement) Model of Labor Force Activity: Extended Tables of Central Tendency, Variation, and Probability Intervals. *Journal of Legal Economics*, 11(1), 23–87.

_________. 2002. Probability Mass Functions for Labor Market Activity Induced by the Markov (Increment-Decrement) Model of Labor Force Activity. *Economics Letters*, 77(3), 425–431.

________. 2009. An Autoregressive Model of Order Two for Worklife Expectancies and Other Labor Force Characteristics with an Application to Major League Baseball. paper presented at the ASSA meeting San Francisco CA.

Endnotes

1. The AR(2) model of years of labor market activity was introduced in the paper "Generalization of the Increment-Decrement Model" by Skoog, presented at the National Association of Forensic Economics sessions of the Allied Social Science Association meeting in January 2003. In that paper, Skoog presented worklife expectancy recursions, and he specified recursions defining probability mass functions for years of activity within an AR(2) model with beginning of period transitions. Skoog and Ciecka presented the paper "An Autoregressive Model of Order Two for Worklife Expectancies and Other Labor Force Characteristics with an Application to Major League Baseball" at the National Association of Forensic Economics sessions of the Allied Social Science Association meeting in January 2009. That paper was an earlier draft of the present paper in which recursions defining probability mass functions within the AR(2) model with beginning of period, mid-period, and end of period transitions were presented. Cushing and Rosenbaum presented a paper "Higher Order Markov Estimates of Worklife: Comparing SIPP to CPS Results" at the National Association of Forensic Economics sessions of the Western Economics Association meeting in 2011 in which they estimated worklife expectancies based on first, second, and third order models. They provided separate worklife expectancies for men and for women, regardless of educational attainment.

2. See Skoog and Ciecka (2001a and 2002) for recursive formulae defining pmfs within the context of the AR(1), Markov, model.

3. One data source classifies hitters in its "Batter Register" (T*he ESPN Baseball Encyclopedia*). Another source summarizes hitters' statistics under the heading "Batting" (*Major League Handbook*). Hitters include position players and, in the American League, designated hitters. Pitchers are excluded, even if they occasionally pinch-hit.

4. Inequality (1a) does not hold throughout our data set. Our data are very thin at older ages and, at times, ${}^{ia}p_x^a > {}^{aa}p_x^a$ at older ages. For example, there is only one player who was inactive at age 37 and active at age 38. That player remained active at age 39; and, therefore, ${}^{ia}p_{38}^a =$ 1 before adjustment for mortality. There are typically only one or two players who were inactive at age $x - 1$ but active at age x for $x \geq 36$ (see Appendix Table 1). Rather than using extreme transition probabilities (often 0 or 1) for ${}^{ia}p_x^a$ for $x \geq 36$, we used ${}^{ia}p_x^a = .9\,{}^{ia}p_{x-1}^a$ for $x \geq 36$ in Table 1. Consider age 42 as another example; in this case ${}^{a}p_{42}^a > {}^{aa}p_{42}^a$. There were 6 players who were active at age 42, three remained active

at age 43; thus ${}^{a}p^{a}_{42} = 3/6 = .5$ before mortality adjustment. There were only 5 players who were active at ages 41 and 42, and two remained active at age 43 leading to ${}^{aa}p^{a}_{42} = 2/5 = .40$ before mortality adjustment. The single hitter who was inactive at age 41 but active at 42 remained active at age 43; thus ${}^{ia}p^{a}_{42} = 1$ prior to accounting for mortality.

5. We assume that ${}^{aa}p^{a}_{x} = {}^{ia}p^{a}_{x} = {}^{ai}p^{a}_{x} = {}^{ii}p^{a}_{x} = {}^{a}p^{a}_{x} = {}^{i}p^{a}_{x} = 0$ for $x \geq 46$, *i.e.*, major league hitters either retire from the game or die at age 46 or beyond. One player did play beyond age 46, but our data set contains few observations for ages older than 39 for initial states *aa* and *ai* and above age 34 for the *ia* category. When exceptions occur to inequality (1), and inequalities (2)-(5) as well, they are at ages younger than 24 or greater than 39.

6. The section *Relations Among Expectancies* in Section IV below proves the first and last inequalities in (3) in the populations and in a sample, with the middle inequality holding empirically in our data but not necessarily generally in the population or a sample.

7. When this is done, using the estimated transition probabilities conditional on survival in Appendix Table 1, starting active at ages 23 through 33 the AR(1) restrictions are rejected at conventional significance levels. Starting inactive, they are also rejected at these ages except for ages 24 and 29.

8. The alternative would be to use capital letters for the states conditional on survival, and to add an additional symbol ($\sim$) on top of the transition probabilities to refer to its estimator, *e.g.*, ${}^{AI}\tilde{p}^{A}_{x}$ would refer to the estimate of those transitioning from activity followed by inactivity at age x into the active state, conditional on survival, for those at age x. We re-interpret ${}^{ai}p^{a}_{x}$ to be this estimator.

9. We use the term *pseudo-recursion* since the right hand side of ${}^{a}\bar{e}^{a}_{x} = .5 + .5{}^{a}p^{a}_{x} + {}^{a}p^{a}_{x}\,{}^{aa}e^{a}_{x+1} + {}^{a}p^{i}_{x}\,{}^{ai}e^{a}_{x+1}$ is not the same as the right hand side of AR(1)-1.

Appendix

Appendix Table 1. Counts of Players by Age within AR(2) Model

Age	${}^{aa}N_x$	${}^{aa}N_x^a$	${}^{aa}N_x^i$	${}^{ia}N_x$	${}^{ia}N_x^a$	${}^{ia}N_x^i$	${}^{ai}N_x$	${}^{ai}N_x^a$	${}^{ai}N_x^i$	${}^{ai}N_x^i$	${}^{ii}N_x^a$	${}^{ii}N_x^i$
19	0	0	0	2	0	2	0	0	0	0	0	0
20	0	0	0	23	21	2	2	1	1	0	0	0
21	20	19	1	61	47	14	2	1	1	1	1	0
22	76	68	8	91	72	19	14	8	6	1	1	0
23	141	126	15	156	127	29	26	14	12	5	0	5
24	258	230	28	158	115	43	44	19	25	13	4	9
25	344	311	33	144	97	47	74	28	46	30	4	26
26	399	341	58	112	66	46	86	25	61	68	6	62
27	401	344	57	84	49	35	104	26	78	103	7	96
28	413	361	52	72	33	39	87	17	70	135	6	129
29	397	333	64	43	15	28	89	12	77	145	14	131
30	350	299	51	43	20	23	96	22	74	152	9	143
31	304	263	41	34	14	20	71	14	57	157	10	147
32	287	240	47	32	14	18	59	12	47	145	7	138
33	261	207	54	22	11	11	62	10	52	116	1	115
34	229	182	47	13	7	6	60	5	55	106	2	104
35	183	140	43	7	3	4	53	3	50	116	0	116
36	141	98	43	2	0	2	45	5	40	125	0	125
37	91	62	29	4	1	3	42	2	40	113	0	113
38	62	39	23	1	1	0	29	2	27	102	0	102
39	42	26	16	2	1	1	22	1	21	81	0	81
40	25	14	11	1	0	1	14	1	13	62	0	62
41	11	6	5	1	1	0	13	1	12	40	0	40
42	5	2	3	1	1	0	6	0	6	28	0	28
43	3	2	1	0	0	0	3	0	3	20	0	20
44	2	1	1	0	0	0	1	0	1	13	0	13
45	1	1	0	0	0	0	1	0	1	8	0	8
46	1	1	0	0	0	0	0	0	0	3	0	3
47	1	1	0	0	0	0	0	0	0	1	0	1
48	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0