MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2023-041 l November 2023
https://doi.org/10.4054/MPIDR-WP-2023-041

Statistical Inference for Discrete-Time Multistate Models: Asymptotic Covariance Matrices, Partial Age Ranges, and Group Contrasts

Daniel C. Schneider l schneider@demogr.mpg.de

This working paper has been approved for release by: Mikko Myrskylä (office-myrskyla@demogr.mpg.de),
Head of the Laboratory of Population Health, the Laboratory of Fertility and Well-Being, and the Research Group Labor Demography.

Statistical Inference for Discrete-Time Multistate Models: Asymptotic Covariance Matrices, Partial Age Ranges, and Group Contrasts

Daniel C. Schneider

Max Planck Institute for Demographic Research, Rostock, Germany

2023-11-02

Abstract

This paper lays out several new asymptotic inference results for discrete-time multistate models. First, it derives asymptotic covariance matrices for the outcome statistics of conditional and/or state expectancies, mean age at first entry, and lifetime risk. It then discusses group comparisons of these outcome measures, which require the calculation of a joint covariance matrix of two or more results. Finally, new procedures are presented for the estimation of multistate models over a partial age range, and how these subrange calculations relate to the result that is obtained from the full age range of the model. All newly derived expressions are compared against bootstrap results in order to verify correctness of results and to assess performance.

1 INTRODUCTION

Discrete-time multistate models (DTMS) have increasingly been used in the social science, demographic, and public health literature.[1] Used as a form of life-course analysis, they posit the lifetime of an individual as a succession of discrete states, the set of which comprises the "state-space". Based on estimated age-specific transition probabilities, the main outcome statistic of these models are state expectancies (expected sojourn times, in Markov chain theory parlance). They differ from continuous-time multistate models, which are amply applied, for example, in the medical literature, by assuming that state transitions occur only in accordance with a fixed time grid. Due to that assumption, research based on DTMS frequently uses longitudinal data on individuals that are sampled at (roughly) regular intervals. Studies using this method are numerous, and they have been applied to a variety of substantive areas. A list of example applications can be found in the introductory section of Dudel and Schneider (2021).

All DTMS calculations take place under the Markov assumption, which posits that transition probabilities do not depend on any previous state visits other than the current one. One-step transition probabilities are the key magnitudes around which other calculations are centered. A suitable method for obtaining (estimating) transition probabilities is to base them on regressions, among which multinomial logistic regression is a popular choice since Millimet *et al.* (2003). Once transition probabilities have been predicted from the regression results, in a third step, formulas from standard Markov chain theory yield state and overall expectancies. A variety of other statistics and refinements have been developed (see, for example, Hunter and Caswell, 2005; Roth and Caswell, 2018; Dudel, 2021). Readers who are completely new to the method are referred to the appendix of Schneider, Myrskylä and van Raalte (2023), which provides an accessible exposition of the basic calculations. Further elaborations in this document assume basic familiarity with this material.

When estimating state expectancies from data, it is of obvious interest to get a measure of how accurately these point estimates are. Most of the applied papers resort to bootstrap techniques for statistical inference, due to the fact that asymptotic formulas have not been derived. There are two exceptions to this. First, Lièvre, Brouard and Heathcote (2003) partially derive asymptotic expressions for variances and covariances of state expectancies. The main limitations of that paper are a) that formulas are not fully explicit, so that the derivates required by the delta method have to be approximated numerically and b) they are derived within a framework (interpolated Markov chains, IMaCh) which imposes many practical limits on the underlying regression. A second exception is Lynch and Brown (2005), who cast the problem in a Bayesian framework, which allows for statements using credible intervals. Here, the main problem lies, as is often the case with Bayesian analysis, in the computational cost of the procedure.

The first contribution of this paper is to fill this gap and derive an expression for the asymptotic covariance matrix of state expectancies that is numerically efficient. All that is necessary to arrive at this result is a set of transition probabilities along with their covariance matrix. As an illustration of this computational requirement, section 2.6 lists (known) formulas of how a covariance matrix of transition probabilities can be obtained from a multinomial logistic regression model, but it must be emphasized that this is not the only method that is suitable. Subsequent calculations are, in principle, independent of how the transition probabilities and their covariance matrix have been obtained. One noteworthy implication of this is that the calculations of this paper are in accordance with complex survey design, which can simply be taken into account at the regression stage. Subsequent steps simply use this adjusted covariance matrix of the regression parameters and hence contain the relevant sampling information for the data.

[1] Other names and abbreviations for the same modelling technique found in the literature are, for example, DTMM, discrete-time multistate life tables (discrete-time MSLTs), or age-stage Markov models.

There are a number of additional contributions in this paper. Section 5 develops asymptotic formulas for a second outcome statistic, mean age at first entry. Section 6 develops formulas that allow for group comparisons, the basis of which are joint covariance matrices for two or more sets of transition probabilities. The only requirement is that all sets of transition probabilities are obtained from the same regression model. Finally, section 7 develops state expectancy results for partial age ranges. The results for partial age ranges are such that, when summing the results from a partition of the full age range, a) point estimates add up to what would have been obtained for the estimate of the entire age range, and b) they also add up in the statistical sense, in that a simple linear combination of the partial age ranges over the partition yields a covariance matrix that would have been obtained directly if the full age range had been used.

2 Preliminary Remarks and Notation

2.1 Ordering and Timing Conventions

Since multistate models in the social sciences are typically cast in terms of the aging of subjects, this document uses the terminology "age" rather than the broader notion of "time", but this is just a naming convention. Whenever "age" is mentioned, it can be read as "time". Similarly, the cross-sectional units need not be human subjects but can be, e.g., firms.

2.1.1 Age-within-Stage, Stage-within-Age, and the ji-Convention

Markov transition matrices can either be written as row-stochastic matrices (rows sum to one) or column-stochastic matrices (columns sum to one). In this document, we use the latter convention.

In the context of multistate modelling, transition matrices can be ordered "age-within-stage" or "stage-within-age". For example, for a model that contains $\bar{s}$ transient states and $\bar{a}$ ages, we define the matrix $\boldsymbol{U}$ as the Markov transition matrix among transient states only:

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{U}_2 & \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} \\ 0 & \boldsymbol{U}_3 & \boldsymbol{0} & & \boldsymbol{0} \\ \vdots & & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{U}_{\bar{a}-1} & \boldsymbol{0} \end{bmatrix} \qquad \text{(stage-within-age)} \qquad (1)$$

The ordering here, which will be used exclusively throughout the text, is "stage-within-age": Each submatrix $\boldsymbol{U}_a$ is $\bar{s} \times \bar{s}$ and has the transition probabilities among the transient states for a particular age z_a. The alternative ordering ("age-within-stage") is

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} & \cdots & \boldsymbol{U}_{1\bar{s}} \\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22} & & \boldsymbol{U}_{2\bar{s}} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{U}_{\bar{s}1} & \boldsymbol{U}_{\bar{s}2} & \cdots & \boldsymbol{U}_{\bar{s}\bar{s}} \end{bmatrix} \qquad \text{(age-within-stage)} \qquad (2)$$

where $\boldsymbol{U}_{ij}$ now is a $\bar{a}_{-1} \times \bar{a}_{-1}$ matrix:

$$\boldsymbol{U}_{ij} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ p_{ij,2} & 0 & & 0 & 0 \\ 0 & p_{ij,3} & & 0 & 0 \\ \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & p_{ij,\bar{a}_{-1}} & 0 \end{bmatrix}$$

As a general rule, *i* is a state index that always refers to the destination state. Index *j* refers to the origin state or to the initial state. Equation (2) follows the convention to denote matrix row and column indexes by *i* and *j*, respectively, and that double subscripts refer first to rows and then to columns. We call this order of subscripts the *ij*-convention or *ij*-notation. Symbols in this text, however, sometimes follow what we call the *ji*-convention if it allows for a more convenient ordering of elements. Then the first subscript refers to the origin state and the second subscript to the destination state. Which convention is used is always stated in the text.

2.1.2 Timing in Markov-related table-like matrices

There are $\bar{a}$ ages in the model and $\bar{a}-1$ age intervals (contained in the vector $\boldsymbol{z}$, in the notation of this article, see below). Irregular age intervals are allowed (but probably difficult to reconcile with estimation from the data). The first age is called the *base age*. The last age is called the *exit age*. At the exit age, all subjects are assumed to enter the absorbing state. Corresponding transition probabilities are set to one.

This can be illustrated in a matrix that contains all relevant information on transition probabilities in a table-like format. A similar matrix is defined in section 2.6 more formally. The structure is illustrated by the following example numbers for transition probabilities:

$$\grave{\boldsymbol{P}} = \begin{matrix} & & p_{11} & p_{12} & p_{13} & p_{21} & p_{22} & p_{23} \\ & & -- & -- & -- & -- & -- & -- \\ 50 & | & . & . & . & . & . & . \\ 60 & | & 0.95 & 0.04 & 0.01 & 0.35 & 0.61 & 0.04 \\ 70 & | & 0.93 & 0.06 & 0.01 & 0.25 & 0.68 & 0.06 \\ 80 & | & 0.86 & 0.10 & 0.04 & 0.15 & 0.74 & 0.12 \\ 90 & | & 0.67 & 0.18 & 0.16 & 0.06 & 0.69 & 0.25 \\ 100 & | & . & . & 1 & . & . & 1 \end{matrix}$$

where orange labels indicate the meaning of columns and rows. The subscripts of the transition probabilities are in ji-format. A dot in the matrix indicates that that matrix element is never used in any calculations. In this example, there are three states. States 1 and 2 are transient and state 3 is absorbing. The model contains ages $\boldsymbol{z} = [50, 60, \dots, 100]$ and $\bar{a} = 6$. 50 is the base age and 100 the exit age. All quantities are stochastic (estimated) except for the entries in the rows corresponding to the base age and the exit age. In the example, the row 3, column 2 number of 0.06 indicates the probability of being in state 2 at age 70, given state 1 at age 60. When exactly this transition takes place is, at this point, still undefined.

2.1.3 Timing in the Data Set and in the Regression

The data set that is used to estimate transition probabilities, e.g., by means of a multinomial logistic model, needs to contain information on the current state and age of subjects. The age value should denote the exact age at observation of the state, not the age at transition. Estimation requires, except for special cases, that the data are equally spaced. In the above example, observations are (should be) roughly ten years apart. Since the unit of observation is a transition, one also needs an (independent) variable that records the origin state. This can just be the lagged dependent variable. The transition probabilities predicted from such a multinomial regression are to be interpreted as the probability that a subject is in state i (not: enters state i) at exact (!) age z_a (see below for notation), given that that subject was in state *j* ten years earlier. We call the ages at which such probabilities are predicted the *prediction ages*. These are the ages contained in $\boldsymbol{z}$ and used to label the matrix $\grave{\boldsymbol{P}}$.

2.1.4 Prediction Ages, Transition Ages, and Standard Transition Timings

It was noted above that the ages in $\boldsymbol{z}$ are prediction ages. The *transition ages*, by contrast, are generally unknown, so an assumption about them must be made. Based on the example model setup from above, Box 1 displays transition timing schemes for several different assumptions. The top section of the box (scheme MID) assumes mid-period transitions, which take place at ages $55, 65, \ldots, 95$. Summing the age intervals shows that a maximum of 45 years can be obtained for the remaining life expectancy. An alternative transition timing assumption is end-of-period, displayed in the second section of the box (scheme EOP). Here the maximum attainable remaining life expectancy is 50 years. Finally, the opposite of the end-of-period assumption is beginning-of-period (section 3, scheme BOP) with a maximum attainable remaining life expectancy of 40 years. For BOP, the base age and the first transition age are identical, indicated by the double vertical bars. One can define more complicated transition timings than MID, EOP, and BOP, but this is a more advanced topic. This theme is pursued in Schneider, Myrskylä and van Raalte (2023).

It is often helpful to think in intervals that are defined by end points determined by both prediction ages and transition ages. For mid-period, this corresponds to the bottom section of Box 1 (scheme MID-SPLIT). For the current example, the length of all subintervals is 5 years. Thinking in these subintervals is helpful for understanding the workings of partial age ranges treated in section 7.

Box 1. Transition Timing Schemes for 10-Year Prediction Intervals from Age 50 to Age 100

Scheme MID

Prediction age	[50]		—		60		—		70		—		80		—		90		—		[100]	
Age at transition	50		55				65				75				85				95			
Length of age interval		5			10				10				10				10					Σ=45

Scheme EOP

Prediction age	[50]		—		60		—		70		—		80		—		90		—		[100]	
Age at transition	50				60				70				80				90				100	
Length of age interval			10				10				10				10				10			Σ=50

Scheme BOP

Prediction age	[50]		—		60		—		70		—		80		—		90		—		[100]	
Age at transition	50				60				70				80				90					
Length of age interval			10				10				10				10							Σ=40

Scheme MID-SPLIT

Prediction age	[50]		—		60		—		70		—		80		—		90		—		[100]	
Age at transition	50		55				65				75				85				95			
Length of age interval		5		5		5		5		5		5		5		5		5				Σ=45

Notes: Different transition timing schemes pertaining to a model with base age 50 and exit age 100 and 10-year prediction intervals in between. Abbreviations used are MID for mid-period, EOP for end-of-period, and BOP for beginning of period. Prediction ages whose transition probabilities are not estimated are placed in brackets.

Two important points from the above should be remembered:

- What one specifies as model ages are prediction ages. They need to be distinguished from transition ages. The latter are unknown and must be pinned down via an assumption.
- When defining a model with $\bar{a}$ ages, only $\bar{a} - 2$ ages are relevant for estimating transition probabilities (see section 2.1.2). They are not estimated for the base and exit ages: The base age of the model is not a proper transition, and the exit age has absorption probabilities of 1. Box 1 shows prediction ages whose transition probabilities are not estimated in brackets.

2.2 NOTATION AND ABBREVIATIONS

The two different types of results treated in this document, life expectancy and mean age at first entry, are abbreviated as LEXP and MAFN, respectively. Each one of LEXP and MAFN results comprise multiple numbers. Lifetime risk, a (single) component of MAFN results, is abbreviated as LRSK. These abbreviations are usually used with upper case letters in order to make text more legible, but sometimes occur in lower case letters, e.g., in graph labeling.

The transition timing assumptions beginning-of-period, mid-period, and end-of-period, which were introduced in section 2.1.4, are sometimes abbreviated as BOP, MID, and EOP, respectively.

All estimated and derived magnitudes in this document follow the normal distribution asymptotically. For notational simplicity, the "hat" notation for estimated magnitudes is suppressed, except for section 2.3.

Unless otherwise specified, vectors are row vectors. If they are given with a subscript, it denotes the number of columns. By contrast, result vectors, like the estimated coefficients of the multinomial logistic estimation, or the vector of state expectancies, are column vectors.

Bar accents indicate matrices that have been summed over the rows and hence are row vectors. A tilde accent indicates that a matrix contains a subset of the information of another matrix. Double-dot accents are used for matrices that combine information for two or more results.

The following table lists the mathematical symbols used in this document. It is meant primarily as a reference table, as most symbols are explicitly introduced and explained in the text.

Symbol	Meaning	Size
a	age index, $a = 1, \ldots, \bar{a}$	
$\bar{a}$	number of ages in the model	
z_a	age corresponding to age index a	
z_1	minimum (baseline) age in model. z_1 is the age of the initial (beginning) state. z_2 is the age at which the first transition takes place.	
$z_{\bar{a}}$	exact age at which all subjects die if they are still alive	
$\boldsymbol{z}, \boldsymbol{z_{-1}}, \boldsymbol{z_{+1}}$	$\boldsymbol{z}$ collects the z_a into a vector; $\boldsymbol{z_{-1}}$ denotes $\boldsymbol{z}$ exlusive of the last element; $\boldsymbol{z_{+1}}$ denotes $\boldsymbol{z}$ exclusive of its first element	$1 \times \bar{a}$ or $1 \times \bar{a}_{-1}$
$\bar{a}_{-1}, \bar{a}_{-2}$	shorthands for $\bar{a} - 1, \bar{a} - 2$	

Symbol	Meaning	Size
n_a	length of age interval starting at age z_a	
$\boldsymbol{n}$	collects the n_a into a vector	$1 \times \bar{a}_{-1}$
s	state index, $s = 1, \dots, \bar{s}$, or $s = 1, \dots, \bar{s}_{+1}$	
$\bar{s}$	number of transient states	
$\bar{s}_{+1}$	shorthand for $\bar{s} + 1$; all derivations are done for a single absorbing state, so $\bar{s}_{+1}$ denotes the number of states in the model	
i, j	state indexes used for transitions; j refers to the origin state or to the initial state; and i refers to the destination state	
h	number of coefficients per mlogit equation	
k	generic index used in different contexts	
$\boldsymbol{P}, p_{ij}$	transition matrix among all states (never actually used in this text); elements are p_{ij}	$\bar{s}\bar{a}_{-1} + 1 \times \bar{s}\bar{a}_{-1} + 1$
$\grave{\boldsymbol{P}}$	matrix that collects the nonzero elements of $\boldsymbol{P}$ in a table-like format	$\bar{a}_{-2} \times \bar{s}\bar{s}_{+1}$
$\boldsymbol{U}, p_{ij}$	submatrix of $\boldsymbol{P}$ consisting of transient states only; since it is a submatrix of $\boldsymbol{P}$, the elements of $\boldsymbol{U}$ are also denoted by p_{ij}.	$\bar{s}\bar{a}_{-1} \times \bar{s}\bar{a}_{-1}$
$\boldsymbol{U_a}$	elements of $\boldsymbol{U}$ pertaining to age a	$\bar{s} \times \bar{s}$
$\boldsymbol{\mathcal{U}}$	collection of the $\boldsymbol{U}_a$, $a = 2, \dots, \bar{a}_{-1}$, in a block row vector	$\bar{s} \times \bar{s}\bar{a}_{-2}$
$\boldsymbol{F}$	fundamental matrix	$\bar{s}\bar{a}_{-1} \times \bar{s}\bar{a}_{-1}$
$\boldsymbol{r}, \boldsymbol{r}_{+1}$	time rewards vector for standard time rewards, such as mid-period; $\boldsymbol{r}_{+1}$ is $\boldsymbol{r}$ exclusive of its first element	$1 \times \bar{a}_{-1}, 1 \times \bar{a}_{-2}$
$\boldsymbol{g}, g_j$	g_j is the fraction of population in state j at baseline age; $\boldsymbol{g}$ collects the g_j in a vector	$1 \times \bar{s}$
$\boldsymbol{I}_c$	identity matrix	$c \times c$
$\boldsymbol{0}, \boldsymbol{1}$	a row vector (one subscript) or matrix (two subscripts) of zeroes or ones. Subscripts indicate the size.	
$\boldsymbol{\mathcal{K}}_{\cdots}^{\cdots}$	3-index permutation matrix	
$\otimes$	Kronecker product	
$\odot$	Hadamard product (elementwise multiplication)	
$\oslash$	elementwise division	

Symbol	Meaning	Size
LEXP-related		
$\boldsymbol{F}_1$	first block column of $\boldsymbol{F}$, with blocks arranged to a block row vector	$\bar{s} \times \bar{s}\bar{a}_{-1}$
$\boldsymbol{E}$	$\boldsymbol{E}$ collects the conditional state expectancies in a matrix of state expectancies (rows), conditional on the initial state (columns)	$\bar{s} \times \bar{s}$
$\boldsymbol{e}^{cond}$	vector of remaining life expectancies, conditional on the initial state j	$\bar{s} \times 1$
$\boldsymbol{e}^{state}$	vector of state expectancies	$\bar{s} \times 1$
e	remaining life expectancy at base age	1×1
$\boldsymbol{E}^{full}, \boldsymbol{e}^{full}$	Matrix and vector containing all elements of $\boldsymbol{E}$, $\boldsymbol{e}^{cond}$, $\boldsymbol{e}^{state}$, and e	$\bar{s}_{+1} \otimes \bar{s}_{+1},\ \bar{s}_{+1}^2 \times 1$
MAFN-related		
$\bar{s}_A$	number of initial and intermediate states	
$\bar{s}_B$	number of target states	
$\bar{s}_{ini}$	number of initial states	
$\boldsymbol{F}_{BA}$	$\boldsymbol{U_a}$ matrices divided up and multiplied according to initial and intermediate states ($\boldsymbol{A}_a$ matrices) and summed target states ($\overline{\boldsymbol{B}}_a$ matrices)	$1 \times \bar{s}_A\bar{a}_{-1}$
$\boldsymbol{g}^{ini}$	initial proportions vector for MAFN initial states only	$1 \times \bar{s}_{ini}$
$\boldsymbol{g}^{im}$	initial proportions vector, modified for MAFN initial and intermediate states	$1 \times \bar{s}$
$\boldsymbol{l}^{raw}$	probabilities of entering the set of target states, by age	$\bar{a}_{-1} \times 1$
l^{lrsk}	lifetime risk	1×1
$\boldsymbol{l}^{nrm}$	normalized probabilities	$\bar{a}_{-1} \times 1$
l^{mafn}	mean age at first entry	1×1
$\boldsymbol{l}^{full}$	vectors $\boldsymbol{l}^{raw}$-l^{mafn} combined	$2\bar{a}_{-1} + 2 \times 1$
Covariance matrices for: (the expression in column "Size" applies to both matrix rows and matrix columns)		
$\boldsymbol{V}^{ml}$	mlogit coefficients	$\bar{s}_{+1}h$
$\boldsymbol{V}^{tr}$	transition probabilities	$\bar{s}\bar{s}_{+1}\bar{a}_{-2}$
$\widetilde{\boldsymbol{V}}^{tr}$	As $\boldsymbol{V}^{tr}$, but with information on transitions to the absorbing state removed	$\bar{s}^2\bar{a}_{-2}$
$\boldsymbol{V}^{F}$	vec $\boldsymbol{F}$	$(\bar{s}\bar{a}_{-1})^2$

Symbol	Meaning	Size
$\boldsymbol{V}^{F_1}$	vec $\boldsymbol{F}_1$	$\bar{s}^2\bar{a}_{-2}$
$\boldsymbol{V}^{E}$	vec $\boldsymbol{E}$	$\bar{s}^2$
$\boldsymbol{V}^{cond}$	$\boldsymbol{e}^{cond}$	$\bar{s}$
$\boldsymbol{V}^{state}$	$\boldsymbol{e}^{state}$	$\bar{s}$
V^{e}	e	1
$\boldsymbol{V}^{full}$	In the context of LEXP: joint covariance matrix for conditional state expectancies, conditional expectancies, state expectancies, and the overall life expectancy	$\bar{s}^2 + 2\bar{s} + 1$
$\boldsymbol{V}^{F_{BA}}$	vec $\boldsymbol{F}_{BA}$	$\bar{s}_A\bar{a}_{-1}$
$\boldsymbol{V}^{raw}$	$\boldsymbol{l}^{raw}$	$\bar{a}_{-1}$
V^{lrsk}	l^{lrsk}	1
$\boldsymbol{V}^{nrm}$	$\boldsymbol{l}^{nrm}$	$\bar{a}_{-1}$
V^{mafn}	l^{mafn}	1
$\boldsymbol{V}^{full}$	In the context of MAFN: joint covariance matrix for raw probabilities, lifetime risk, normalized probabilities, and mean age at first entry (i.e., for $\boldsymbol{l}^{full}$)	$2\bar{a}_{-1} + 2$

Matrices used in the transformation of one covariance matrix into another (delta method)

$\boldsymbol{G}^{tr}$	$\boldsymbol{V}^{ml} \rightarrow \boldsymbol{V}^{tr}$	$\bar{s}\bar{s}_{+1}\bar{a}_{-2} \times \bar{s}_{+1}h$
$\boldsymbol{G}^{F_1}$	$\widetilde{\boldsymbol{V}}^{tr} \rightarrow \boldsymbol{V}^{F_1}$	$\bar{s}^2\bar{a}_{-2} \times \bar{s}^2\bar{a}_{-2}$
$\boldsymbol{G}^{F_{BA}}$	$\boldsymbol{V}^{tr} \rightarrow \boldsymbol{V}^{F_{BA}}$	$\bar{s}\bar{a}_{-1} \times \bar{a}_{-1}(\bar{s}_A^2 + \bar{s}_B)$
$\boldsymbol{G}^{nrm}$	$\boldsymbol{V}^{raw} \rightarrow \boldsymbol{V}^{nrm}$	$\bar{a}_{-1} \times \bar{a}_{-1}$
$\boldsymbol{G}^{full}$	LEXP: $\boldsymbol{V}^{E} \rightarrow \boldsymbol{V}^{full}$	$\bar{s}^2 + 2\bar{s} + 1 \times \bar{s}^2$
	MAFN: $\boldsymbol{V}^{raw} \rightarrow \boldsymbol{V}^{full}$	$2\bar{a}_{-1} + 2 \times \bar{a}_{-1}$

Matrices that hold information for two or more results

$\ddot{\boldsymbol{V}}^{Fx}$	joint covariance matrix of the elements of any two matrices out of $\boldsymbol{F}_1, \boldsymbol{F}_{BA}$	*varies*
$\ddot{\boldsymbol{V}}^{comb}$	joint covariance matrix of the elements of any two matrices out of $\boldsymbol{E}, \boldsymbol{l}^{raw}$	*varies*
$\ddot{\boldsymbol{V}}^{full}$	joint covariance matrix of the elements of any two matrices out of $\boldsymbol{E}^{full}$, $\boldsymbol{l}^{full}$	*varies*

Symbol	Meaning	Size
$\ddot{\boldsymbol{G}}^{Fx}$	$\ddot{\boldsymbol{V}}^{tr} \rightarrow \ddot{\boldsymbol{V}}^{Fx}$	$varies$
$\ddot{\boldsymbol{G}}^{full}$	$\ddot{\boldsymbol{V}}^{comb} \rightarrow \ddot{\boldsymbol{V}}^{full}$	$varies$

Partial age ranges

b_k	age index of beginning age
e_k	age index of ending age

2.3 Asymptotic Covariance Matrices and the Delta-Method

The delta method provides an approximation of the covariance matrix of a nonlinear function of a parameter vector. The central results of this paper are based on repeated application of this technique: The distribution of the transition probabilities is usually obtained by applying the delta method to the covariance matrix of the regression parameters. The distribution of state expectancies (LEXP), in turn, is obtained by applying the delta method to the covariance matrix of the transition probabilities. For MAFN results, the delta method is applied even twice: Once for the raw probabilities, and once more for the normalized probabilities (see section 5.2).

The delta method is connected to asymptotic analysis in that it (only) requires asymptotic normality, and in particular, consistency, as far as properties of an estimator are concerned. Therefore, derivations take as their point of departure an estimator that is root-N consistent and asymptotically normal. A large variety of standard regression methods possess this property, which is established using a law of large numbers and a central limit theorem, under mild assumptions concerning the regression error terms. For a parameter vector $\boldsymbol{\beta}_N$ whose dependence on the sample size N is, for the purpose of exposition in this section only, explicitly indicated, this is expressed as

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\right) \overset{d}{\rightarrow} Normal(\mathbf{0}, \boldsymbol{V}^{\infty}) \tag{3}$$

where $\boldsymbol{V}^{\infty}$ is the asymptotic covariance matrix of $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\right)$. Read the d-arrow as "converges in distribution to..." as N goes to infinity. An alternative notation is $\overset{a}{\sim}$, which is to be read as "is asymptotically distributed as...". As the sample size goes to infinity, the covariance matrix of a consistent estimator of $\boldsymbol{\beta}$ itself goes to zero, of course, but the multiplication by the rate of convergence $\sqrt{N}$ in (3) ensures convergence in distribution.[2] For practical purposes, the covariance matrix of $\widehat{\boldsymbol{\beta}}_N$ is then estimated to be $\widehat{\boldsymbol{V}}^{\beta} = \frac{\widehat{\boldsymbol{V}}^{\infty}}{N}$. Formulas for $\widehat{\boldsymbol{V}}^{\beta}$ are available for many regression methods. For example, for ordinary linear squares (OLS) regression under the homoskedasticity assumption, $\widehat{\boldsymbol{V}}^{\beta} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, where $\boldsymbol{X}$ is the data (regressor) matrix and $\hat{\sigma}^2$ the standard error of the regression.

This paper will be concerned with matrices whose elements are estimated from data. The covariance matrix of such a matrix is then expressed using the vec operator that stacks the columns of a matrix, so for a generic $m \times n$ matrix $\boldsymbol{A}$, the covariance matrix $\boldsymbol{V}^{A} = \text{cov}(\text{vec}\,\boldsymbol{A})$ is of size $mn \times mn$. A standard result from normal distribution theory is that if vec $\boldsymbol{A}$ is (exactly) distributed as $Normal(\mathbf{0}, \boldsymbol{V}^{A})$, then, for a conformable matrix $\boldsymbol{G}$ the linear mapping $\mathbf{GA}$ is distributed

[2] $N^c\left(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\right)$ converges to zero in probability for $0 \leq c < 0.5$.

as $Normal(\mathbf{0}, \boldsymbol{GVG}')$. It can be shown that this result also holds in asymptotic analysis. But what about nonlinear mappings? Denote by $\boldsymbol{\mathcal{G}}$ a (matrix-valued) nonlinear differentiable mapping. If $\boldsymbol{B} = \boldsymbol{\mathcal{G}}(\boldsymbol{A})$, we define

$$\boldsymbol{G}^B = \frac{\partial \operatorname{vec} \boldsymbol{\mathcal{G}}(\boldsymbol{A})}{\partial \operatorname{vec}(\boldsymbol{A})'} \tag{4}$$

The delta method establishes that, if (3) holds for $\operatorname{vec} \widehat{\boldsymbol{A}}_N$, (4) leads to

$$\sqrt{N}\left(\operatorname{vec} \boldsymbol{\mathcal{G}}(\widehat{\boldsymbol{A}}_N) - \operatorname{vec} \boldsymbol{\mathcal{G}}(\boldsymbol{A})\right) \xrightarrow{d} Normal\left(\mathbf{0}, \boldsymbol{G}^B \boldsymbol{V}^\infty \boldsymbol{G}^{B\prime}\right) \tag{5}$$

To give an example, let $\boldsymbol{B} = \boldsymbol{A}^2$. Since $\boldsymbol{G}^B = \frac{\partial \operatorname{vec} \boldsymbol{A}^2}{\partial \operatorname{vec}(\boldsymbol{A})'} = \boldsymbol{A}' \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{A}$ (see, for example, Lütkepohl 1996, p.189), where $\boldsymbol{I}$ denotes the identity matrix, we get $\sqrt{N}\left(\operatorname{vec} \widehat{\boldsymbol{A}}^2 - \operatorname{vec} \boldsymbol{A}^2\right) \xrightarrow{d} Normal(0, (\boldsymbol{A}' \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{A})\boldsymbol{V}^\infty(\boldsymbol{A}' \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{A})')$. For inference on $\widehat{\boldsymbol{A}}$, we use $\widehat{\boldsymbol{V}}^A = \frac{\widehat{\boldsymbol{V}}^\infty}{N}$. For inference on $\widehat{\boldsymbol{B}} = \boldsymbol{\mathcal{G}}(\widehat{\boldsymbol{A}})$, we use $\widehat{\boldsymbol{G}}^B \widehat{\boldsymbol{V}}^A \widehat{\boldsymbol{G}}^{B\prime}$. We write $\widehat{\boldsymbol{G}}^B$ in "hat" notation to emphasize that the derivative is to be evaluated at the parameter estimates.

The reasoning behind the delta method is based on a Taylor series approximation (or a mean value expansion) and the consistency of the estimator. For more information on the delta method, see, for example, section 5.6 of Davidson and MacKinnon (2004) or section 3.5 of Wooldridge (2002).

The notation used in this paper is simplified throughout in that, in addition to the omission of the "hat" notation, the explicit dependence of magnitudes on the sample size is suppressed. It is implicit that all arguments are made within an asymptotic (not exact) framework.

2.4 3-Index Orderings

We will encounter more complex (3-index) orderings than in (2) when discussing covariance matrices. In addition to being ordered according to the origin or initial state *j* and destination state *i*, matrices have a sort order with respect to the age index *a*. The sort order is indicated by an index triplet, with the first index being the slowest moving index and the last index being the fastest moving index. For example, a sort order of a-j-i indicates a sort order by age, then by origin or initial state, and then by destination state.

Reordering operations will sometimes be necessary, and the below outlines how this can be achieved in mathematical formulas. Key ingredient are commutation matrices (see, e.g., Lütkepohl, 2005, p. 662). A commutation matrix converts, for an $m \times n$ matrix $\boldsymbol{A}$, between $\operatorname{vec}(\boldsymbol{A})$ and $\operatorname{vec}(\boldsymbol{A}')$:

$$\operatorname{vec}(\boldsymbol{A}') = \boldsymbol{K}_{mn} \operatorname{vec}(\boldsymbol{A}) \tag{6}$$

$\boldsymbol{K}$ is a permutation matrix and has the properties $\boldsymbol{K}'_{mn} = \boldsymbol{K}^{-1}_{mn} = \boldsymbol{K}_{nm}$, so we also have

$$\operatorname{vec}(\boldsymbol{A}) = \boldsymbol{K}_{nm} \operatorname{vec}(\boldsymbol{A}') \tag{7}$$

From (6), premultiplying a matrix by $\boldsymbol{K}_{mn}$ reorders its rows from sort order n-m to m-n. Taking the transpose,

$$\operatorname{vec}(\boldsymbol{A}')' = \operatorname{vec}(\boldsymbol{A})' \boldsymbol{K}'_{mn} = \operatorname{vec}(\boldsymbol{A})' \boldsymbol{K}_{nm} \tag{8}$$

we see that the same reordering operations on the columns of a matrix are achieved by postmultiplying by the transpose of $\boldsymbol{K}_{mn}$.

Reordering operations can be extended to 3-index orderings. For example, $\boldsymbol{K}_{a(ji)}$, which uses the product of two index limits, resorts j-i-$a \rightarrow a$-j-i. By extension, using $\boldsymbol{K}_{i(aj)}$ and $\boldsymbol{K}_{j(ia)}$, this implies a full circular set of possible reorderings of j-i-$a \rightarrow a$-j-$i \rightarrow i$-a-$j \rightarrow j$-i-a, so repeated sort operations can switch from any of those orderings to any other. What this does not cover, however, are reorderings that leave the left or the right index unchanged, e.g., j-i-$a \rightarrow j$-a-i. It is easy to see that the required permutation matrix here is $\boldsymbol{I}_{\bar{s}} \otimes \boldsymbol{K}_{ai}$ ($\bar{s}$ denotes the number of elements over which the index j runs): $\boldsymbol{K}_{ai}$ resorts i-$a \rightarrow a$-i, and this is repeated within each level of j, which yields the desired operation. j-a-i, in turn, then can be reordered j-a-$i \rightarrow i$-j-$a \rightarrow a$-i-$j \rightarrow j$-a-i (using $\boldsymbol{K}_{i(ja)}$, $\boldsymbol{K}_{a(ij)}$, and $\boldsymbol{K}_{j(ai)}$). We have covered all six possible orderings, and again, repeated ordering operations can switch from any of these six orderings to any other. Such operations will be summarized by a matrix $\boldsymbol{\mathcal{K}}$ with a subcript and superscript that indicate the new and original orderings, respectively. For example, premultiplication of a matrix by $\boldsymbol{\mathcal{K}}_{ija}^{aji}$ reorders the rows of that matrix a-j-$i \rightarrow i$-j-a.

Note that such matrices implicitly assume that the correct index bounds are used. For example, $\boldsymbol{\mathcal{K}}_{ija}^{aji}$ may operate on a matrix with dimension $\bar{s}_{+1}\bar{s}\bar{a}$ or on a matrix with dimension $\bar{s}\bar{s}\bar{a}_{-2}$ (where $\bar{s}_{+1}$ and $\bar{a}_{-2}$ are shorthands for $\bar{s}+1$ and $\bar{a}-2$, respectively; see the notation section). It is implicitly assumed that $\boldsymbol{\mathcal{K}}_{ija}^{aji}$ has been constructed correctly in each case, even if multiplications using differently sized matrices happen in the same formula.

2.5 Example Data Set and Application, Bootstrap, and Replication Code

Several sections in this document compare results based on the new asymptotic derivations to results obtained from the simulation of life histories or to bootstrap results. The same data set and the same example application are used throughout. They are described in this section. The data set is not a real-world data set, but a simulated one. It is taken from the Stata package "dtms" (Schneider, 2023), which is publicly available for download. By using freely available data, readers can follow the replication code (see below) more easily and reproduce or modify estimation results more easily.

The simulated data set concerns fictitious annual longitudinal survey data containing life histories with respect to cognitive impairment, starting at age 50. It consists of the following variables:

id	subject ID
n	subject observation number
year	survey year
age	exact age at interview (years, centered at 50)
cog3	cognitive impairment, 3-category
sex	sex, 2-category
educ	education level, 3-category
numdrinks	number of alcoholic drinks per week

The data are used to fit a multinomial logistic regression with cog3 as the dependent variable, which records the four states of the multistate model: no/mild/severe impairment (transient states), and dead (absorbing state). The numeric encoding of states that is sometimes used is 1-4, in the preceding order of states (1=no impairment through 4=dead). All possible transitions from any transient state to any other transient state or to the absorbing state occur in the data.

Independent variables are cog3 lagged by 1 period, linear and quadratic age, a full two-way interaction between sex and educ, and, without interaction and as a continuous variable, numdrinks. No random or fixed effects are in the model, but standard errors are clustered by subject. Using the fitted model, different sets of transition probabilities are predicted at fixed values of the independent variables, setting these to either the sample average or to specific categories. Prediction ages begin age 50, increase by single age year, and end at age 110. In a last step, the results treated in this paper (LEXP and MAFN) are calculated from the transition probabilities.

The resulting confidence intervals (CIs) are compared to CIs obtained via a nonparametric bootstrap with 500 replications. Both bootstrap CIs that are based on the implied standard error (SEs) of the coefficients as well as bootstrap percentile intervals are presented.

The calculation of both LEXP and MAFN requires the weighting of numbers that are conditional on the initial state, using the initial proportion of states in the sample as weights. For simplicity, this paper applies a fixed proportion of initial states (88% without impairment, 10% mildly impaired, 2% severely impaired) that holds true for the overall sample also to subgroups (e.g., women). Since derivations in this paper are conditional on the initial proportions and do not take into account the uncertainty from estimating them, applying a single fixed proportion to all subgroups does not influence the comparison of asymptotic CIs against bootstrap CIs.

A script that replicates all calculated results this paper accompanies this article and can be accessed under https://osf.io/nxeaf. All calculations were performed using the Stata package "dtms" (Schneider, 2023) in Stata 18.

2.6 Caveats

Covariances as derived in this document are conditional on the initial proportions vector $\boldsymbol{g}$. This is a shortcoming, since initial proportions are oftentimes estimated from the data. For large samples, however, $\boldsymbol{g}$ is estimated precisely, so ignoring its variation should lead to a negligible bias only. As a further caveat, in the derivations below, covariances are conditional on fixed, non-stochastic time rewards. This includes the standard transition timing cases beginning-of-period, mid-period, and end-of-period.

All formulas are developed for a single absorbing state but this is inconsequential for the statistics in this paper. When estimating state expectancies, which concern transient states only, the multiple absorbing states can just be lumped into a single one at the regression stage. When estimating mean age at first entry, the difference between transient and absorbing states does not matter, since one can specify either transient states or the absorbing state as the target state (see section 5).

3 Transition Probabilities

As was mentioned in the introduction, this section illustrates just one out of many ways of how transition probabilities and their covariance matrix can be obtained. It uses known formulas for multinomial logistic regression. They are shown here explicitly in order to give a comprehensive account of relevant calculations for all steps of the process, but could be replaced by formulas for a different regression method, for example.

Given an $\bar{s}_{+1}h \times 1$ estimated multinomial logistic coefficient vector $\boldsymbol{\beta}^{ml}$, the transition probabilites are calculated as:

$$p_{j1a} = \Pr\left(i = 1|\overline{\boldsymbol{x}}_{ja}\right) = \frac{1}{1 + \sum_{i=2}^{\bar{s}+1} \exp\left(\overline{\boldsymbol{x}}_{ja}\boldsymbol{\beta}_i^{ml}\right)} \tag{9}$$

$$p_{jsa} = \Pr(i = s|\overline{\boldsymbol{x}}_{ja}) = \frac{\exp(\overline{\boldsymbol{x}}_{ja}\boldsymbol{\beta}_s^{ml})}{1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}_{ja}\boldsymbol{\beta}_i^{ml})} \tag{10}$$

where p_{jia} is the transition probability from state j to state i at age a and $\overline{\boldsymbol{x}}_{ja}$ contains the appropriate prediction values for origin state and age and the means of the other covariates. The above assumes, without loss of generality, that $s = 1$ is the base outcome.

The covariance matrix of the transition probabilities is calculated according to the delta method. We define (according the ji-convention) $\boldsymbol{p}_{ji} = (p_{ji2} \dots p_{ji\bar{a}_{-1}})'$, and the $\bar{a}_{-2} \times \bar{s}\bar{s}_{+1}$ matrix

$$\grave{\boldsymbol{P}} = [\boldsymbol{p}_{11} \dots \boldsymbol{p}_{1\bar{s}_{+1}} \dots \boldsymbol{p}_{\bar{s}1} \dots \boldsymbol{p}_{\bar{s}\bar{s}_{+1}}] \tag{11}$$

Since $\boldsymbol{\beta}^{ml}$ is asymptotically normally distributed with associated asymptotic covariance matrix $\boldsymbol{V}_{ml}$ and since $\grave{\boldsymbol{P}} = \mathcal{G}(\boldsymbol{\beta}^{ml})$ is a differentiable mapping of $\boldsymbol{\beta}^{ml}$, we obtain the asymptotic covariance matrix for $\text{vec}(\grave{\boldsymbol{P}})$, denoted by $\boldsymbol{V}^{tr}$, as

$$\boldsymbol{V}^{tr} = \boldsymbol{G}^{tr}\boldsymbol{V}^{ml}\boldsymbol{G}^{tr\prime} \tag{12}$$

using

$$\boldsymbol{G}^{tr} = \frac{\partial \, \text{vec} \, \grave{\boldsymbol{P}}}{\partial \boldsymbol{\beta}^{ml\prime}} \tag{13}$$

The derivation of $\boldsymbol{G}^{tr}$ uses the quotient rule for derivatives, which for the scalar case reads

$$\frac{\partial \left(\frac{u(x)}{v(x)}\right)}{\partial x} = \frac{vu' - uv'}{v^2}$$

Using $\overline{\boldsymbol{x}}$ as a shorthand for $\overline{\boldsymbol{x}}_{jk}$ and $\boldsymbol{\beta}_i$ for $\boldsymbol{\beta}_i^{ml}$, we get for outcome 1:

$$\frac{\partial p_{j1k}}{\partial \boldsymbol{\beta}'} = \frac{\left(1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_i)\right) * \boldsymbol{0}_{\bar{s}_{+1}h} - 1 * [0, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_2), \dots, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_{\bar{s}+1})] \otimes \overline{\boldsymbol{x}}}{\left(1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_i)\right)^2}$$

$$= -\frac{[0, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_2), \dots, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_{\bar{s}+1})] \otimes \overline{\boldsymbol{x}}}{\left(1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_i)\right)^2}$$

The derivative for outcomes $s > 1$ is:

$$\frac{\partial p_{jsk}}{\partial \boldsymbol{\beta}'} = \frac{\left(1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_i)\right) \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_s)\overline{\boldsymbol{x}}_s - \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_s)[0, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_2), \dots, \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_{\bar{s}+1})] \otimes \overline{\boldsymbol{x}}}{\left(1 + \sum_{i=2}^{\bar{s}+1} \exp(\overline{\boldsymbol{x}}\boldsymbol{\beta}_i)\right)^2}$$

with $\overline{\boldsymbol{x}}_s$ being a $1 \times \bar{s}_{+1}h$ vector that contains $\overline{\boldsymbol{x}}$ for the entries corresponding to the s^{th} equation and is zero otherwise.

The ordering of elements in $\boldsymbol{V}^{tr}$ is as in $\text{vec} \, \grave{\boldsymbol{P}}$, i.e., j-i-a. $\boldsymbol{V}^{tr}$ contains covariance information for transitions into the absorbing state. For some purposes below, this information needs to be removed to make matrices conformable. To do so, we first reorder to i-j-a, define a cut-off matrix $\boldsymbol{C} = [\boldsymbol{I}_{\bar{s}^2\bar{a}_{-2}} \quad \boldsymbol{0}_{\bar{s}^2\bar{a}_{-2},\bar{s}\bar{a}_{-2}}]$ whose purpose it is to remove the last $\bar{s}\bar{a}_{-2}$ rows of a matrix, and define

$$\widetilde{\boldsymbol{V}}^{tr} = \boldsymbol{\mathcal{K}}_{jia}^{ija} * \boldsymbol{C} * \boldsymbol{\mathcal{K}}_{ija}^{jia} * \boldsymbol{V}^{tr} * \boldsymbol{\mathcal{K}}_{ija}^{jia\prime} * \boldsymbol{C}' * \boldsymbol{\mathcal{K}}_{jia}^{ija\prime} \tag{14}$$

$\widetilde{\boldsymbol{V}}^{tr}$ has the same ordering as $\boldsymbol{V}^{tr}$, but any information on transitions to the absorbing state has been removed.

4 LIFE EXPECTANCY

4.1 POINT ESTIMATES

First, we can define a fixed "rewards" row vector $\boldsymbol{r}$ for regular transition timing schemes, based on the length of age intervals (n_a):

$$\boldsymbol{r} = \begin{cases} [0, n_1, n_2, \ldots, n_{\bar{a}-2}] & \text{(beginning-of-period)} \\ [n_1 + 0, n_2 + n_1, \ldots, n_{\bar{a}-1} + n_{\bar{a}-2}]/2 & \text{(mid-period)} \\ [n_1, n_2, \ldots, n_{\bar{a}-1}] & \text{(end-of-period)} \end{cases} \tag{15}$$

The LEXP point estimates are calculated via selected elements of the fundamental matrix

$$\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{U})^{-1} \tag{16}$$

where $\boldsymbol{U}$ contains the estimated transition probabilities among the transient states and is ordered "stage-within-age" (see (1)). F can be written as

$$\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{U})^{-1} = \sum_{a=0}^{\bar{a}-2} \boldsymbol{U}^a = \left[\begin{array}{c|cccc} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{U}_2 & \boldsymbol{I} & & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{U}_3 \cdot \boldsymbol{U}_2 & \boldsymbol{U}_3 & & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{U}_{\bar{a}-1} \cdot \ldots \cdot \boldsymbol{U}_2 & \boldsymbol{U}_{\bar{a}-1} \cdot \ldots \cdot \boldsymbol{U}_3 & \cdots & \boldsymbol{U}_{\bar{a}-1} & \boldsymbol{I} \end{array}\right]$$

The elements of the first block column, which in the above equation are separated by a vertical line and highlighted in red, are the ones that indicate probabilities of reaching states at later ages, given a particular state at the baseline age. These entries are necessary for calculating the life expectancy at the base age, which typically is the quantity of interest. Reorder those elements as a block-row vector

$$\boldsymbol{F}_1 = \left[\boldsymbol{I} \quad \boldsymbol{U}_2 \quad \boldsymbol{U}_3\boldsymbol{U}_2 \quad \cdots \quad \prod_{a=\bar{a}_{-1}}^{2} \boldsymbol{U}_a\right] \tag{17}$$

The only thing that needs to be done to obtain life expectancy numbers is to weigh the probabilities of reaching certain states according to the length of the age intervals and the transition timing. This is done by

$$\boldsymbol{E} = \boldsymbol{F}_1(\boldsymbol{r}' \otimes \boldsymbol{I}_{\bar{s}}) \tag{18}$$

which results in a $\bar{s} \times \bar{s}$ matrix. Simple and weighted sums of the elements of $\boldsymbol{E}$ lead to conditional expectancies, state expectancies, and overall remaining life expectancy:

$$\boldsymbol{e}^{cond} = (\boldsymbol{1}_{\bar{s}} * \boldsymbol{E})' \tag{19}$$

$$\boldsymbol{e}^{state} = \boldsymbol{E} * \boldsymbol{g}' \tag{20}$$

$$e = \boldsymbol{1}_{\bar{s}} * \boldsymbol{e}^{state} \tag{21}$$

It suggests itself to present the point estimates in a $\bar{s}_{+1} \times \bar{s}_{+1}$ matrix

$$\boldsymbol{E}^{full} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{e}^{state} \\ \boldsymbol{e}^{cond\,\prime} & e \end{bmatrix} \tag{22}$$

since the last row and last column of that matrix are the sum (rows) and weighted sum (columns) of the preceding rows and columns, respectively. For the purpose of this document, however, sometimes a slightly different ordering is needed than the vectorization of the above matrix, we define a column vector of length $\bar{s}^2 + 2\bar{s} + 1 = \bar{s}_{+1}^2$:

$$\boldsymbol{e}^{full} = \begin{bmatrix} \operatorname{vec} \boldsymbol{E} \\ \boldsymbol{e}^{cond} \\ \boldsymbol{e}^{state} \\ e \end{bmatrix} \tag{23}$$

For an alternative exposition of the above, see section 1.1.2 of the appendix of Schneider, Myrskylä and van Raalte (2023).

4.2 COVARIANCE MATRICES

To obtain LEXP standard errors, a naïve approach would first calculate $\boldsymbol{V}^F$, the full covariance matrix of all elements of $\boldsymbol{F}$. Which, as the appendix section 10.1 shows, is

$$\boldsymbol{V}^F = (\boldsymbol{F}' \otimes \boldsymbol{F})\boldsymbol{V}^U(\boldsymbol{F}' \otimes \boldsymbol{F})' \tag{24}$$

One problem with this approach is that in section 3 we have calculated $\boldsymbol{V}^{tr}$, which is the covariance matrix of the nonzero elements of $\boldsymbol{U}$. $\boldsymbol{V}^U$ could be obtained by reordering elements and inserting rows and columns of zeroes into $\boldsymbol{V}^{tr}$. However, there is a second, more serious problem: The size of these matrices grows so fast that for plausible real-world model setups, computation would be infeasible.

A better approach therefore is to only calculate the covariance matrix for the elements of $\boldsymbol{F}$ that are needed for life expectancy at the base age, that is, for the elements contained in $\boldsymbol{F}_1$. We redefine it slightly as

$$\boldsymbol{F}_1 = \left[\boldsymbol{U}_2 \quad \boldsymbol{U}_3\boldsymbol{U}_2 \quad \cdots \quad \prod_{a=\bar{a}_{-1}}^{2} \boldsymbol{U}_a\right] = [\boldsymbol{f}_2 \quad \boldsymbol{f}_3 \quad \cdots \quad \boldsymbol{f}_{\bar{a}_{-1}}] \tag{25}$$

where we have removed the first block element of (17), the identity matrix, which does not contribute to stochastic variation. We have also introduced $\boldsymbol{f}_a$ as a shorthand for the matrix products of the $\boldsymbol{U}_a$.

Our goal is to apply the delta method for $\boldsymbol{F}_1 = \boldsymbol{g}(\grave{\boldsymbol{P}})$ in order to obtain

$$\boldsymbol{V}^{F_1} = \boldsymbol{G}^{F_1}\widetilde{\boldsymbol{V}}^{tr}\boldsymbol{G}^{F_1}{}' \tag{26}$$

The elements of $\grave{\boldsymbol{P}}$ that $\boldsymbol{F}_1$ contains are

$$\boldsymbol{\mathcal{U}} = [\boldsymbol{U}_2 \quad \boldsymbol{U}_3 \quad \cdots \quad \boldsymbol{U}_{\bar{a}_{-1}}]$$

so we seek

$$\boldsymbol{G}^{F_1} = (\dots)\frac{\partial \operatorname{vec} \boldsymbol{F}_1}{\partial \operatorname{vec}(\boldsymbol{\mathcal{U}})'}(\dots)'$$

where the ellipses indicate potential reordering operations, which remain to be made more precise. Recall the matrix derivative result that for generic matrices $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, if only $\boldsymbol{Y}$ depends on the vector $\boldsymbol{c}$,

$$\frac{\partial \operatorname{vec}(\boldsymbol{XYZ})}{\partial \boldsymbol{c}'} = (\boldsymbol{Z}' \otimes \boldsymbol{X})\frac{\partial \operatorname{vec} \boldsymbol{Y}}{\partial \boldsymbol{c}'} \tag{27}$$

(see, e.g., Lütkepohl, 2005, p.668). Applying this result to the above, we get

$$\frac{\partial\,\mathrm{vec}(\boldsymbol{f}_i)}{\partial\,\mathrm{vec}(\boldsymbol{U}_j)'} = \frac{\partial\,\mathrm{vec}\left(\prod_{a=i}^{j+1}\boldsymbol{U}_a * \boldsymbol{U}_j * \prod_{a=j-1}^{2}\boldsymbol{U}_a\right)}{\partial\,\mathrm{vec}(\boldsymbol{U}_j)'}$$

using the convention that the first product term evaluates to $\boldsymbol{I}_{\bar{s}}$ for $j > i$. The below shows the full matrix derivative with rows and columns labeled according to the (unvectorized) expression that is being differentiated (row labels) and the (unvectorized) arguments of differentiation (column labels):

$$\frac{\partial\,\mathrm{vec}\,\boldsymbol{F}_1}{\partial\,\mathrm{vec}(\boldsymbol{\mathcal{U}})'} = \tag{28}$$

	$\boldsymbol{U}_2$	$\boldsymbol{U}_3$	$\boldsymbol{U}_4$	$\cdots$	$\boldsymbol{U}_{\bar{a}_{-1}}$
$\boldsymbol{U}_2$	$\boldsymbol{I}\otimes\boldsymbol{I}$	$\mathbf{0}$	$\mathbf{0}$	$\cdots$	$\mathbf{0}$
$\boldsymbol{U}_3\boldsymbol{U}_2$	$\boldsymbol{I}\otimes\boldsymbol{U}_3$	$\boldsymbol{U}_2{}'\otimes\boldsymbol{I}$	$\mathbf{0}$		$\mathbf{0}$
$\boldsymbol{U}_4\boldsymbol{U}_3\boldsymbol{U}_2$	$\boldsymbol{I}\otimes\boldsymbol{U}_4\boldsymbol{U}_3$	$\boldsymbol{U}_2{}'\otimes\boldsymbol{U}_4$	$(\boldsymbol{U}_3\boldsymbol{U}_2)'\otimes\boldsymbol{I}$		$\mathbf{0}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\boldsymbol{U}_{\bar{a}_{-1}}\ldots\boldsymbol{U}_2$	$\boldsymbol{I}\otimes\boldsymbol{U}_{\bar{a}_{-1}}\ldots\boldsymbol{U}_3$	$\boldsymbol{U}_2{}'\otimes\boldsymbol{U}_{\bar{a}_{-1}}\ldots\boldsymbol{U}_4$	$(\boldsymbol{U}_3\boldsymbol{U}_2)'\otimes\boldsymbol{U}_{\bar{a}_{-1}}\ldots\boldsymbol{U}_5$	$\cdots$	$(\boldsymbol{U}_{\bar{a}_{-2}}\ldots\boldsymbol{U}_2)'\otimes\boldsymbol{I}$

where identity matrices have dimensions $\bar{s}\times\bar{s}$ and zero matrices $\bar{s}^2\times\bar{s}^2$. Reordering elements, we use

$$\boldsymbol{G}^{F_1} = \boldsymbol{\mathcal{K}}_{jia}^{aji}\frac{\partial\,\mathrm{vec}\,\boldsymbol{F}_1}{\partial\,\mathrm{vec}(\boldsymbol{\mathcal{U}})'}\boldsymbol{\mathcal{K}}_{jia}^{aji\prime} \tag{29}$$

to calculate $\boldsymbol{V}^{F_1}$ in (26). We obtain the covariance matrix of conditional state expectancies by applying linear combinations to the appropriate matrix blocks

$$\boldsymbol{V}^E = (\boldsymbol{I}_{\bar{s}^2}\otimes\boldsymbol{r}_{+1})\boldsymbol{V}^{F_1}(\boldsymbol{I}_{\bar{s}^2}\otimes\boldsymbol{r}_{+1})' \tag{30}$$

where $\boldsymbol{r}_{+1}$ is the vector $\boldsymbol{r}$ as defined in (15) exclusive of its first element. This element is omitted since conditional expectancies assume a particular initial state, which therefore does not contribute to stochastic variation. The covariances of conditional expectancies, state expectancies, and the total life expectancy, which are simple linear combinations, are

$$\boldsymbol{V}^{cond} = (\boldsymbol{I}_{\bar{s}}\otimes\mathbf{1}_{\bar{s}})\boldsymbol{V}^E(\boldsymbol{I}_{\bar{s}}\otimes\mathbf{1}_{\bar{s}})' \tag{31}$$

$$\boldsymbol{V}^{state} = (\boldsymbol{g}\otimes\boldsymbol{I}_{\bar{s}})\boldsymbol{V}^E(\boldsymbol{g}\otimes\boldsymbol{I}_{\bar{s}})' \tag{32}$$

$$V^e = (\boldsymbol{g}\otimes\mathbf{1}_{\bar{s}})\boldsymbol{V}^E(\boldsymbol{g}\otimes\mathbf{1}_{\bar{s}})' \tag{33}$$

The combined singular covariance matrix of rank $\bar{s}^2$ that corresponds to (23) is obtained by simply performing all of the above calculations in a single step via

$$\boldsymbol{V}^{full} = \boldsymbol{G}^{full}\boldsymbol{V}^E\boldsymbol{G}^{full\prime} \tag{34}$$

using the full linear combination vector

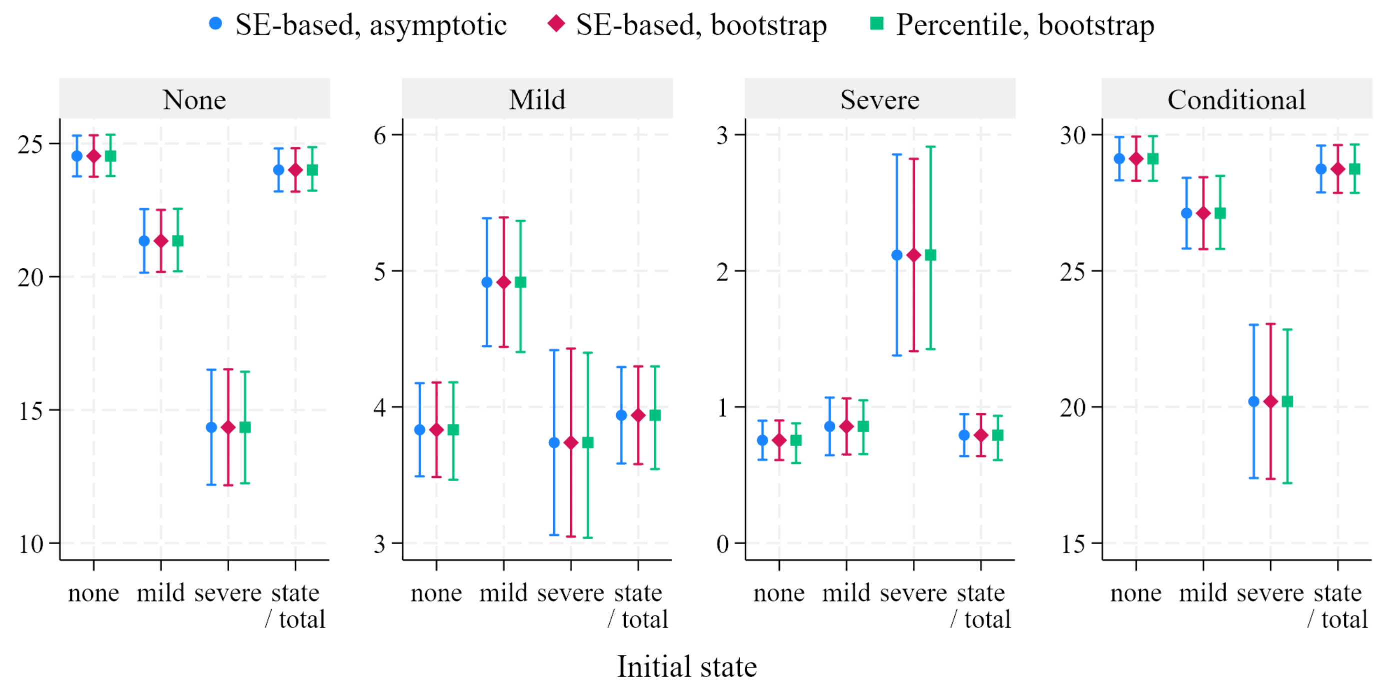
$$\boldsymbol{G}^{full} = \begin{bmatrix}\boldsymbol{I}_{\bar{s}^2}\\ \boldsymbol{I}_{\bar{s}}\otimes\mathbf{1}_{\bar{s}}\\ \boldsymbol{g}\otimes\boldsymbol{I}_{\bar{s}}\\ \boldsymbol{g}\otimes\mathbf{1}_{\bar{s}}\end{bmatrix} \tag{35}$$

4.3 Comparison to Bootstrap Results

The cognitive impairment example with three transient states (see section 2.5) is used for a comparison of asymptotic CIs for LEXP results, calculated according to the derivations from the previous section, with CIs from a nonparametric bootstrap with 500 replications. The comparison is presented in Figure 1, which is divided into four subgraphs. The four plot triplets of each subgraph correspond to one row of matrix (19); that is, the order of the 16 plot triplets, from left to right over the full graph, i.e., across subgraph headings (outcome states) and subgraph categorical axes (initial states), in terms of equation (19), is $\text{vec}\left[\boldsymbol{E}^{full'}\right]$ (note the transpose). This particular ordering and layout of the graph puts all points in a subgraph on roughly the same scale, which facilitates the visual assessment of the agreement of asymptotic and bootstrap CIs.

Two types of bootstrap CIs are shown. The first one is based on the standard errors of coefficient estimates, calculated over all replications. The second one is a percentile interval spanning the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the coefficient distribution over all replications. All point estimates depicted are asymptotic ones, including the ones that are used as the midpoint for bootstrap CIs.[3] All of the estimated magnitudes have asymptotic CIs that are very close to their bootstrap counterparts.

Figure 1: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: State and Overall Life Expectancies



Notes: Subgraphs 1-3 show state expectancies, by initial state (categorical axis: none, mild, or severe) and weighted using initial state proportions. The rightmost subgraph shows conditional life expectancies, by initial state and weighted using initial state proportions. The four plot triplets of each subgraph correspond to one row of matrix (22); that is, the order of the 16 plot triplets, from left to right over the full graph, i.e., across subgraph headings (outcome states) and subgraph categorical axes (initial states), in terms of equation (22), is $\text{vec}\left(\boldsymbol{E}^{full'}\right)$ (note the transpose). Blue dots and whiskers show 95% asymptotic confidence intervals based on the derivations in this article. Red diamonds and whiskers depict 95% confidence intervals based on the standard errors obtained from 500 bootstrap samples; and green squares and whiskers show 95% bootstrap percentile intervals. Each point estimate triplet uses a single value: the asymptotic one.

[3] For a comparison between asymptotic and bootstrap CIs where the bootstrap point estimates are not taken to be the asymptotic ones, see Table 1 in section 6.6.

5 LIFETIME RISK AND MEAN AGE AT FIRST ENTRY

The particular mechanism of mean age at first entry (MAFN) implemented here assumes three different types of states: An initial set of states, an optional intermediate one, and a set of target states. If an intermediate set of states is specified, subjects can move back and forth between initial and intermediate states any number of times (including zero). What is of interest is the first transition to the set of target states. The set of target states can be comprised of any mixture of transient states and/or the absorbing state.

5.1 POINT ESTIMATES

Denote the number of initial and intermediate states by $\bar{s}_A$ and the number of target states by $\bar{s}_B$. Let the $\bar{s}_A \times \bar{s}_A$ matrix $\boldsymbol{A}_a$ denote the age a transitions among the initial and intermediate states only. Let the $\bar{s}_B \times \bar{s}_A$ matrix $\boldsymbol{B}_a$ denote the age a transitions from the initial and intermediate states to the target states. The $1 \times \bar{s}_A$ matrix $\overline{\boldsymbol{B}}_a$ is $\boldsymbol{B}_a$ summed over the rows ($\overline{\boldsymbol{B}}_a = \boldsymbol{1}_{\bar{s}_B} * \boldsymbol{B}_a$). It is helpful to assume, as the following does, that the order among initial and intermediate states in the $\boldsymbol{A}_a$ (columns and rows) and $\overline{\boldsymbol{B}}_a$ (columns) is such that the initial states come first, followed by intermediate states (if any are specified). Define

$$\boldsymbol{F}_{BA} = \left[\overline{\boldsymbol{B}}_2 \quad \overline{\boldsymbol{B}}_3 \boldsymbol{A}_2 \quad \overline{\boldsymbol{B}}_4 \boldsymbol{A}_3 \boldsymbol{A}_2 \quad \cdots \quad \overline{\boldsymbol{B}}_{\bar{a}} \prod_{a=\bar{a}_{-1}}^{2} \boldsymbol{A}_a \right] = [\boldsymbol{f}_2 \quad \boldsymbol{f}_3 \quad \cdots \quad \boldsymbol{f}_{\bar{a}}] \tag{36}$$

where we have redefined the $\boldsymbol{f}_a$, which are now of dimension $1 \times \bar{s}_A$. They contain the probabilities, conditional on the initial state, of moving into the set of target states at ages $2, \ldots, \bar{a}$, given that subjects can only move through initial and intermediate states before arriving at a target state.

We define as $\boldsymbol{g}^{im}$ the initial proportions vector, ordered such that initial states come first and rescaled, if necessary, to sum to one:

$$\boldsymbol{g}^{im} = \left[\boldsymbol{g}^{ini} \quad \boldsymbol{0}_{\bar{s}_{med}}\right] / \left(\boldsymbol{1}_{\bar{s}_{ini}} * \boldsymbol{g}^{ini}\right)$$

where $\boldsymbol{g}^{ini}$ contain the original initial proportions of the initial states and $\bar{s}_{ini}$ and $\bar{s}_{med}$ are the number of initial and intermediate states, respectively. Then the age-specific "raw" probabilities of moving to any one of the target states is

$$\boldsymbol{l}^{raw} = \left(\boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}^{im}\right)\boldsymbol{F}'_{BA} \tag{37}$$

Summing over all elements gives the lifetime risk:

$$l^{rsk} = \boldsymbol{1}_{\bar{a}_{-1}} * \boldsymbol{l}^{raw} \tag{38}$$

Normalized probabilities, which sum to one, are

$$\boldsymbol{l}^{nrm} = \frac{\boldsymbol{l}^{raw}}{l^{rsk}} \tag{39}$$

Mean age at first entry (conditional on ever moving to a target state) is a weighted sum of the normalized probabilities, using as weights the transition ages:

$$l^{mafn} = (\boldsymbol{z}_{-1} + \boldsymbol{n}) * \boldsymbol{l}^{nrm} \tag{40}$$

The above is the formula applicable to end-of-period transitions. For beginning-of-period transitions, replace $\boldsymbol{n}$ by $\boldsymbol{0}$. For mid-period transitions, replace $\boldsymbol{n}$ by $\boldsymbol{n} / 2$. Finally, the full vector of results is

$$\boldsymbol{l}^{full} = \begin{bmatrix} \boldsymbol{l}^{raw} \\ \boldsymbol{l}^{lrsk} \\ \boldsymbol{l}^{nrm} \\ \boldsymbol{l}^{mafn} \end{bmatrix} \tag{41}$$

5.2 COVARIANCE MATRICES

Calculations resemble to some extent those in section 4.2. For calculating

$$\boldsymbol{V}^{F_{BA}} = \boldsymbol{G}^{F_{BA}} \breve{\boldsymbol{V}}^{tr} \boldsymbol{G}^{F_{BA}\prime} \tag{42}$$

we again use the result on matrix differentiation from section 4.2. The difference now is that we have to additionally differentiate with respect to the $\bar{\boldsymbol{B}}_a$. When differentiating with respect to the $\boldsymbol{A}_a$, we get

$$\boldsymbol{G}_A^{F_{BA}} = \tag{43}$$

$$\begin{array}{c|cccccc}
 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \boldsymbol{A}_4 & \cdots & \boldsymbol{A}_{\bar{a}_{-2}} & \boldsymbol{A}_{\bar{a}_{-1}} \\
\hline
\bar{\boldsymbol{B}}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_3\boldsymbol{A}_2 & \boldsymbol{I}\otimes\bar{\boldsymbol{B}}_3 & \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_4\boldsymbol{A}_3\boldsymbol{A}_2 & \boldsymbol{I}\otimes\bar{\boldsymbol{B}}_4\boldsymbol{A}_3 & \boldsymbol{A}_2{}'\otimes\bar{\boldsymbol{B}}_4 & \boldsymbol{0} & & \boldsymbol{0} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_4\boldsymbol{A}_4\boldsymbol{A}_3\boldsymbol{A}_2 & \boldsymbol{I}\otimes\bar{\boldsymbol{B}}_5\boldsymbol{A}_4\boldsymbol{A}_3 & \boldsymbol{A}_2{}'\otimes\bar{\boldsymbol{B}}_5\boldsymbol{A}_4 & (\boldsymbol{A}_3\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_5 & & \boldsymbol{0} & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\
\bar{\boldsymbol{B}}_{\bar{a}_{-1}}\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_2 & \boldsymbol{I}\otimes\bar{\boldsymbol{B}}_{\bar{a}_{-1}}\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_3 & \boldsymbol{A}_2{}'\otimes\bar{\boldsymbol{B}}_{\bar{a}_{-1}}\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_4 & (\boldsymbol{A}_3\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_{\bar{a}_{-1}}\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_5 & \cdots & (\boldsymbol{A}_{\bar{a}_{-3}}\ldots\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_{\bar{a}_{-1}} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_2 & \boldsymbol{I}\otimes\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_3 & \boldsymbol{A}_2{}'\otimes\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_4 & (\boldsymbol{A}_3\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_5 & & (\boldsymbol{A}_{\bar{a}_{-3}}\ldots\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}} & (\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_2)'\otimes\bar{\boldsymbol{B}}_{\bar{a}}
\end{array}$$

where identity matrices are of dimension $\bar{s}_A \times \bar{s}_A$ and zero matrices of dimension $\bar{s}_A \times \bar{s}_A^2$. Differentiating with respect to $\bar{\boldsymbol{B}}_a$ yields

$$\boldsymbol{G}_B^{F_{BA}} = \tag{44}$$

$$\begin{array}{c|cccccc}
 & \bar{\boldsymbol{B}}_2 & \bar{\boldsymbol{B}}_3 & \bar{\boldsymbol{B}}_4 & \cdots & \bar{\boldsymbol{B}}_{\bar{a}_{-1}} & \bar{\boldsymbol{B}}_{\bar{a}} \\
\hline
\bar{\boldsymbol{B}}_2 & \boldsymbol{I}_{\bar{s}_A} & \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_3\boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{A}_2{}' & \boldsymbol{0} & & \boldsymbol{0} & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_4\boldsymbol{A}_3\boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{0} & (\boldsymbol{A}_3\boldsymbol{A}_2)' & & \boldsymbol{0} & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\
\bar{\boldsymbol{B}}_{\bar{a}_{-1}}\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & (\boldsymbol{A}_{\bar{a}_{-2}}\ldots\boldsymbol{A}_2)' & \boldsymbol{0} \\
\bar{\boldsymbol{B}}_{\bar{a}}\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} & (\boldsymbol{A}_{\bar{a}_{-1}}\ldots\boldsymbol{A}_2)'
\end{array}$$

The expressions in $\boldsymbol{G}_B^{F_{BA}}$ do not have Kronecker products because the identity matrix that forms part of them is just the scalar one. We then have

$$\boldsymbol{G}^{F_{BA}} = \begin{bmatrix} \boldsymbol{G}_A^{F_{BA}} & \boldsymbol{G}_B^{F_{BA}} \end{bmatrix} \tag{45}$$

$\breve{\boldsymbol{V}}^{tr}$ in (42) is slightly different from $\boldsymbol{V}^{tr}$ originally defined in (12). The latter matrix is ordered j-i-a. The central derivative in the derivation of the LEXP covariances in section 4.2, $\boldsymbol{G}^{F_1}$ (see (29)), was reordered accordingly. The same procedure is insufficient for (45) because we have divided up the states of the model (represented by matrices $\boldsymbol{U}_a$ in section 4.2) into transitions among initial and intermediate states (matrices $\boldsymbol{A}_a$) and transitions into target states (matrices $\boldsymbol{B}_a$). We therefore have to divide up $\boldsymbol{V}^{tr}$ in an analogous fashion. The order of the rows and columns of $\boldsymbol{V}^{tr}$ must, in a first block, follow a-j-i, where both j and i run over the initial and intermediate states; followed by an ordering of a-j-i, where j runs over the target states and i over the initial and intermediate states. Moreover, for that second part of the ordering, elements have to be summed over j (the target states). This last step accounts for the conversion of the $\boldsymbol{B}_a$ matrices to $\overline{\boldsymbol{B}}_a$. For simplicity, we denote by $\breve{\boldsymbol{V}}^{tr}$ the matrix that embodies the above transformations but do not show corresponding transformation formulas explicitly.

The raw probabilities $\boldsymbol{l}^{raw}$ defined in (37) have the covariance matrix

$$\boldsymbol{V}^{raw} = \left(\boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}^{im}\right)\boldsymbol{V}^{F_{BA}}\left(\boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}^{im}\right)'$$

In order to derive the covariances of the normalized probabilities (39), we have to apply the delta method again. We have

$$\boldsymbol{V}^{nrm} = \boldsymbol{G}^{nrm}\boldsymbol{V}^{raw}\boldsymbol{G}^{nrm\prime}$$

and, recalling (39),

$$\boldsymbol{G}^{nrm} = \frac{\partial \boldsymbol{l}^{nrm}}{\partial \boldsymbol{l}^{raw\prime}} = \frac{\left(\boldsymbol{1}_{\bar{a}_{-1}} * \boldsymbol{l}^{raw} * \boldsymbol{I}_{\bar{a}_{-1}} - \boldsymbol{l}^{raw} * \boldsymbol{1}_{\bar{a}_{-1}}\right)}{\left(\boldsymbol{1}_{\bar{a}_{-1}} * \boldsymbol{l}^{raw}\right)^2} \tag{46}$$

or, in explicit notation,

$$\boldsymbol{G}^{nrm} = \begin{pmatrix} \sum l_{a-1} - l_1 & -l_1 & \cdots & -l_1 \\ -l_2 & \sum l_{a-1} - l_2 & & -l_2 \\ \vdots & & \ddots & \vdots \\ -l_{a-1} & -l_{a-1} & \cdots & \sum l_{a-1} - l_{a-1} \end{pmatrix} \times \frac{1}{\left(\sum l_{a-1}\right)^2} \tag{47}$$

where l_{a-1} denotes the $(a-1)^{th}$ element of $\boldsymbol{l}^{raw}$ and sums are taken over $a = 2, \ldots, \bar{a}$.[4] Remember from (40) that the mean age at first entry is calculated as a linear combination of the transition ages and normalized probabilities, which is the last piece of information necessary to write the covariance matrices of raw probabilities, lifetime risk, normalized probabilities, and mean age at first entry as

$$\boldsymbol{V}^{raw} = \left(\boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}^{im}\right)\boldsymbol{V}^{F_{BA}}\left(\boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}^{im}\right)' \tag{48}$$

$$V^{lrsk} = \boldsymbol{1}_{\bar{a}_{-1}}\boldsymbol{V}^{raw}\boldsymbol{1}'_{\bar{a}_{-1}} \tag{49}$$

$$\boldsymbol{V}^{nrm} = \boldsymbol{G}^{nrm}\boldsymbol{V}^{raw}\boldsymbol{G}^{nrm\prime} \tag{50}$$

$$V^{mafn} = (\boldsymbol{z}_{-1} + \boldsymbol{n})\boldsymbol{V}^{nrm}(\boldsymbol{z}_{-1} + \boldsymbol{n})' \tag{51}$$

Accordingly, the combined singular covariance matrix of rank $\bar{a}_{-1}$ that corresponds to (41) is calculated as

$$\boldsymbol{V}^{full} = \boldsymbol{G}^{full}\boldsymbol{V}^{raw}\boldsymbol{G}^{full\prime} \tag{52}$$

using the full linear combination vector

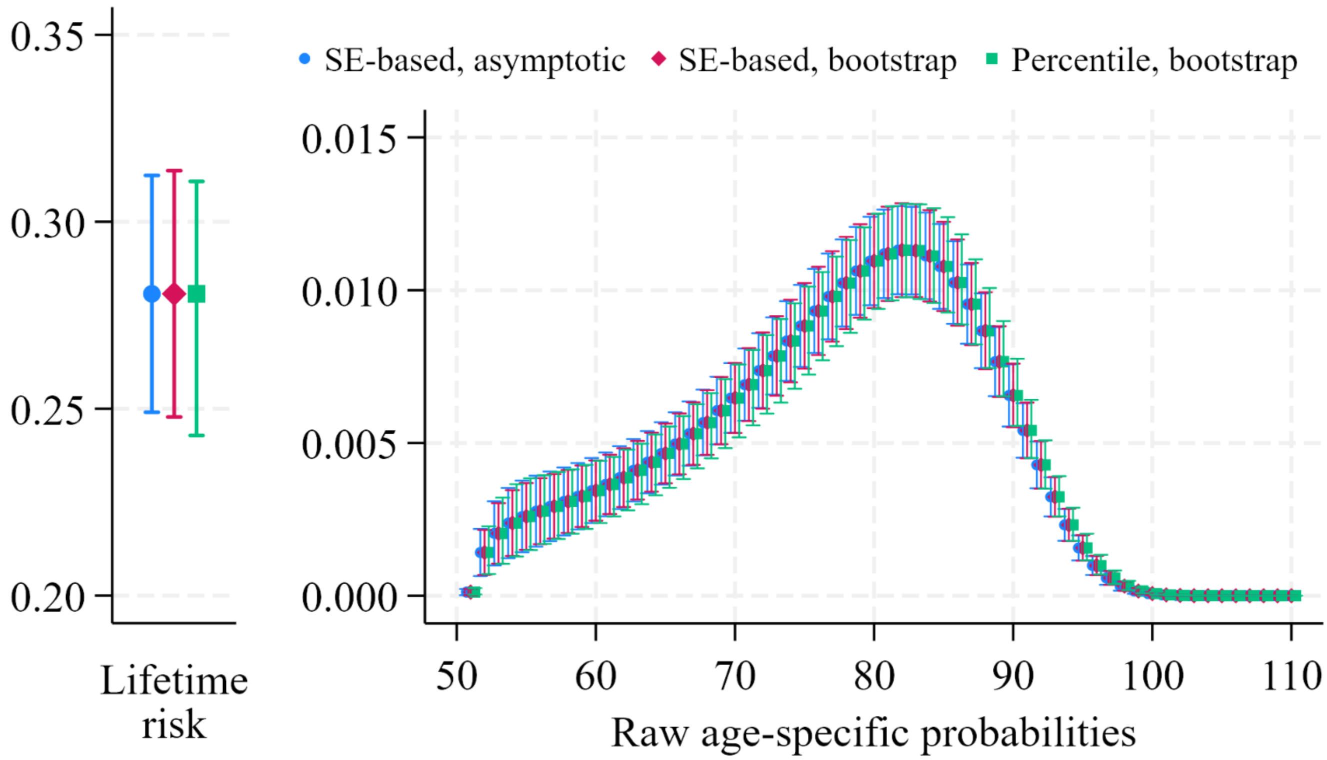
[4] The indexing used may seem a little awkward but is required by the correct usage of the age index a. Timing and age conventions used are such that $a = 1$ corresponds to z_1, the baseline age, which has a zero probability of moving to the target state since, by assumption, the subject is in one of the initial states. At the other end of the age range, at age $z_{\bar{a}}$, the probability of moving to the target state may be nonzero. This is because the target states are allowed to include the absorbing state, which, again by assumption, is transitioned to at age $z_{\bar{a}}$ if the subject is still alive. Consequently, $\boldsymbol{l}^{raw}$ has $\bar{a} - 1$ elements and the sum starts at $a = 2$.

$$\boldsymbol{G}^{full} = \begin{bmatrix} \boldsymbol{I}_{\bar{a}_{-1}} \\ \boldsymbol{1}_{\bar{a}_{-1}} \\ \boldsymbol{G}^{nrm} \\ (\boldsymbol{z}_{-1} + \boldsymbol{n})\boldsymbol{G}^{nrm} \end{bmatrix} \tag{53}$$

5.3 COMPARISON TO BOOTSTRAP RESULTS

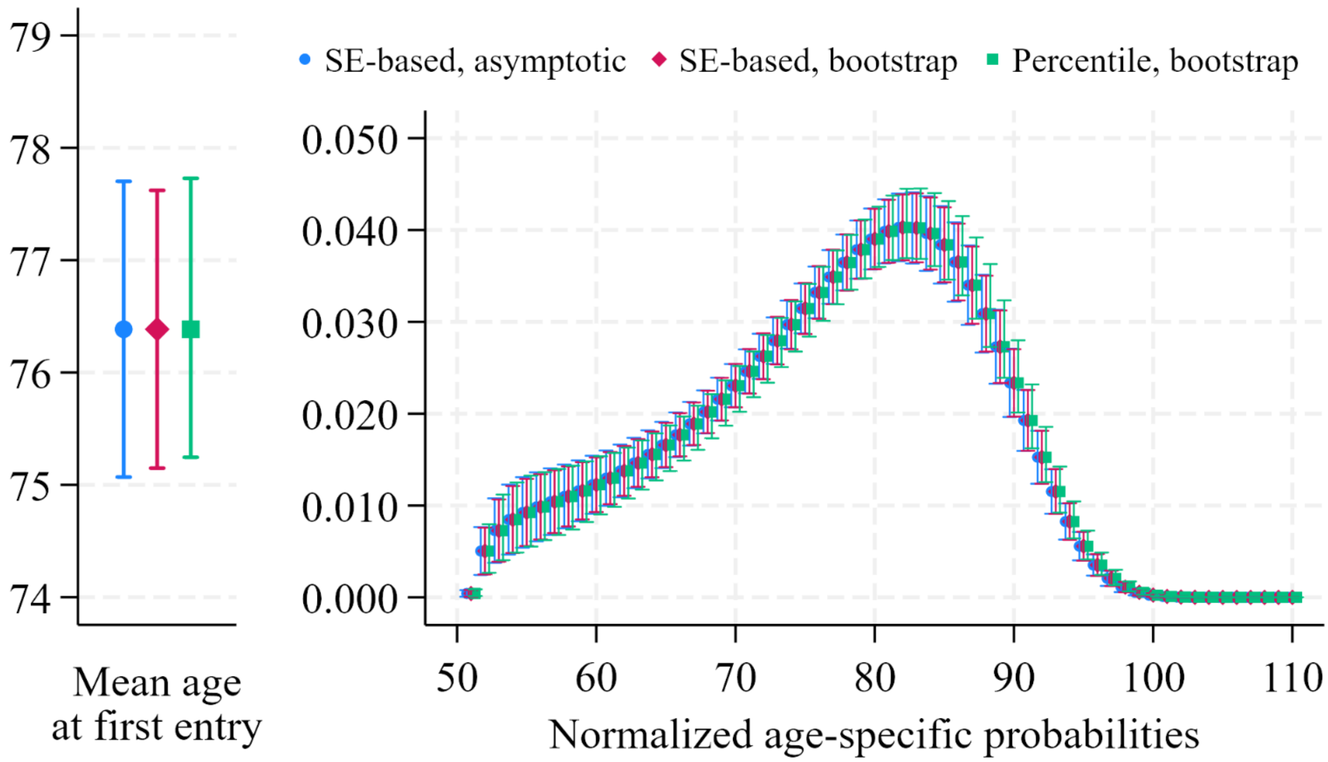
In order to check the quality of the asymptotic approximations, we use MAFN estimates from the cognitive impairment regression where we specify the initial state(s) as unimpaired, the intermediate state(s) as mildly impaired, and the target state(s) as severely impaired. The visual comparison of the resulting asymptotic CIs versus bootstrap CIs is divided into two parts. Figure 2 shows results for lifetime risk (left-hand subgraph) and its component probabilities (right-hand subgraph); and Figure 3 for mean age at first entry and its component probabilities. Both figures show discrepancies between asymptotic and bootstrap results that are larger than for LEXP. However, discrepancies are still relatively small.

Figure 2: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: Lifetime Risk and Raw Age-Specific Probabilities



Notes: The left-hand graph shows lifetime risk, which is the simple sum of raw (unmodified) age-specific probabilities of moving to the set of target states. These are shown on the right-hand graph. Marker symbols, capped lines, and colors are as in Figure 1.

Figure 3: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: Mean Age at First Entry and Normalized Age-Specific Probabilities



Notes: The left-hand graph shows mean age at first entry, which is the age-weighted sum of the normalized age-specific probabilities of moving to the set of target states. These are shown on the right-hand graph. Marker symbols, capped lines, and colors are as in Figure 1.

6 Group Comparisons

In this section, a double-dot accent is used for all matrices that contain information for two or more groups.

6.1 Transition Probabilities

When inferentially comparing two groups, the first step is to generate a joint covariance matrix of the transition probabilities. This is easily achieved by using (14) with

$$\ddot{\boldsymbol{G}}^{tr} = \begin{bmatrix} \boldsymbol{G}_1^{tr} \\ \boldsymbol{G}_2^{tr} \end{bmatrix} \tag{54}$$

to get

$$\ddot{\boldsymbol{V}}^{tr} = \ddot{\boldsymbol{G}}^{tr} \boldsymbol{V}^{ml} \ddot{\boldsymbol{G}}^{tr\prime} \tag{55}$$

$\boldsymbol{G}_1^{tr}$ is as in (13) with $\grave{\boldsymbol{P}}$ being replaced by $\grave{\boldsymbol{P}}_1$, the table of transition probabilities for group 1; and likewise, for $\boldsymbol{G}_2^{tr}$.

In the following, it is assumed that $\ddot{\boldsymbol{V}}^{tr}$ correctly contains (e.g., MAFN) or correctly does not contain (e.g., LEXP) information for transitions to the absorbing state, depending on the context. See the discussion surrounding (14).

6.2 LIFE EXPECTANCY

The combined (joint) covariance matrix for life expectancy is then calculated by the familiar formula (26):

$$\ddot{\boldsymbol{V}}^{F_1} = \ddot{\boldsymbol{G}}^{F_1}\ddot{\boldsymbol{V}}^{tr}\ddot{\boldsymbol{G}}^{F_1\prime} \tag{56}$$

but with

$$\ddot{\boldsymbol{G}}^{F_1} = \begin{bmatrix} \boldsymbol{G}_1^{F_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{F_1} \end{bmatrix} \tag{57}$$

The group-specific matrices $\boldsymbol{G}_1^{F_1}$ and $\boldsymbol{G}_2^{F_1}$ are each calculated as in (28)-(29). The combined covariance matrix corresponding to (30) is

$$\ddot{\boldsymbol{V}}^{E} = \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,2} \end{bmatrix} \ddot{\boldsymbol{V}}^{F_1} \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,2} \end{bmatrix}^{\prime} \tag{58}$$

where $\boldsymbol{r}_{+1,1}$ and $\boldsymbol{r}_{+1,2}$ denote transition timing vectors as defined in section 4.1 for group 1 and group 2, respectively. The full joint covariance matrix corresponding to (34) is

$$\ddot{\boldsymbol{V}}^{full} = \ddot{\boldsymbol{G}}^{full}\ddot{\boldsymbol{V}}^{E}\ddot{\boldsymbol{G}}^{full\prime} \tag{59}$$

with

$$\ddot{\boldsymbol{G}}^{full} = \begin{bmatrix} \boldsymbol{G}_1^{full} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{full} \end{bmatrix} \tag{60}$$

In the construction of $\boldsymbol{G}_1^{full}$ and $\boldsymbol{G}_2^{full}$, see (35), the group-specific initial proportions $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ have been used.

6.3 MEAN AGE AT FIRST ENTRY

In analogy to the calculations in the previous subsection, we extend the single-group formula (42) to

$$\ddot{\boldsymbol{V}}^{F_{BA}} = \ddot{\boldsymbol{G}}^{F_{BA}}\ddot{\boldsymbol{V}}^{tr}\ddot{\boldsymbol{G}}^{F_{BA}\prime} \tag{61}$$

by defining $\ddot{\boldsymbol{G}}^{F_{BA}}$ as

$$\ddot{\boldsymbol{G}}^{F_{BA}} = \begin{bmatrix} \boldsymbol{G}_1^{F_{BA}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{F_{BA}} \end{bmatrix}$$

$\boldsymbol{G}_1^{F_{BA}}$ and $\boldsymbol{G}_2^{F_{BA}}$ consist of (43)-(45) calculated for groups 1 and 2, respectively. $\ddot{\boldsymbol{V}}^{tr}$ must have been reordered for (61) within groups as described in section 5.2. The full joint covariance matrix corresponding to (52) is

$$\ddot{\boldsymbol{V}}^{full} = \ddot{\boldsymbol{G}}^{full}\ddot{\boldsymbol{V}}^{raw}\ddot{\boldsymbol{G}}^{full\prime} \tag{62}$$

In (62), equation (48) is expanded for group comparison as

$$\ddot{\boldsymbol{V}}^{raw} = \begin{bmatrix} \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_1^{im} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_2^{im} \end{bmatrix} \ddot{\boldsymbol{V}}^{F_{BA}} \begin{bmatrix} \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_1^{im} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_2^{im} \end{bmatrix}^{\prime} \tag{63}$$

Moreover, we now have

$$\ddot{\boldsymbol{G}}^{full} = \begin{bmatrix} \boldsymbol{G}_1^{full} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{full} \end{bmatrix}$$

where $\boldsymbol{G}_1^{full}$ and $\boldsymbol{G}_2^{full}$ are constructed as in (53) using the group-specific matrices $\boldsymbol{G}_1^{nrm}$ and $\boldsymbol{G}_2^{nrm}$ (see (46)) as well as group-specific transition ages $(\boldsymbol{z}_{-1} + \boldsymbol{n}_1)$ and $(\boldsymbol{z}_{-1} + \boldsymbol{n}_2)$.

6.4 COMBINING DIFFERENT TYPES OF RESULTS

A combined covariance matrix can be constructed for any mixture of results. For example, one can combine life expectancy and mean age at first entry – potential complications of reordering MAFN-related matrices aside – simply by writing the formulas of the previous subsections as

$$\ddot{\boldsymbol{G}}^{Fx} = \begin{bmatrix} \boldsymbol{G}_1^{F_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{F_{BA}} \end{bmatrix}$$

$$\ddot{\boldsymbol{V}}^{Fx} = \ddot{\boldsymbol{G}}^{Fx} \ddot{\boldsymbol{V}}^{tr} \ddot{\boldsymbol{G}}^{Fx\prime}$$

$$\ddot{\boldsymbol{V}}^{comb} = \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_2^{im} \end{bmatrix} \ddot{\boldsymbol{V}}^{Fx} \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\bar{a}_{-1}} \otimes \boldsymbol{g}_2^{im} \end{bmatrix}'$$

$$\ddot{\boldsymbol{G}}^{full} = \begin{bmatrix} \boldsymbol{G}_1^{full} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{full} \end{bmatrix}$$

$$\ddot{\boldsymbol{V}}^{full} = \ddot{\boldsymbol{G}}^{full} \ddot{\boldsymbol{V}}^{comb} \ddot{\boldsymbol{G}}^{full\prime}$$

where $\ddot{\boldsymbol{V}}^{comb}$ is the covariance matrix of combined/mixed result types. A subscript of 1 in this instance refers to an expression pertaining to life expectancy and a subscript of 2 to an expression for mean age at first entry.

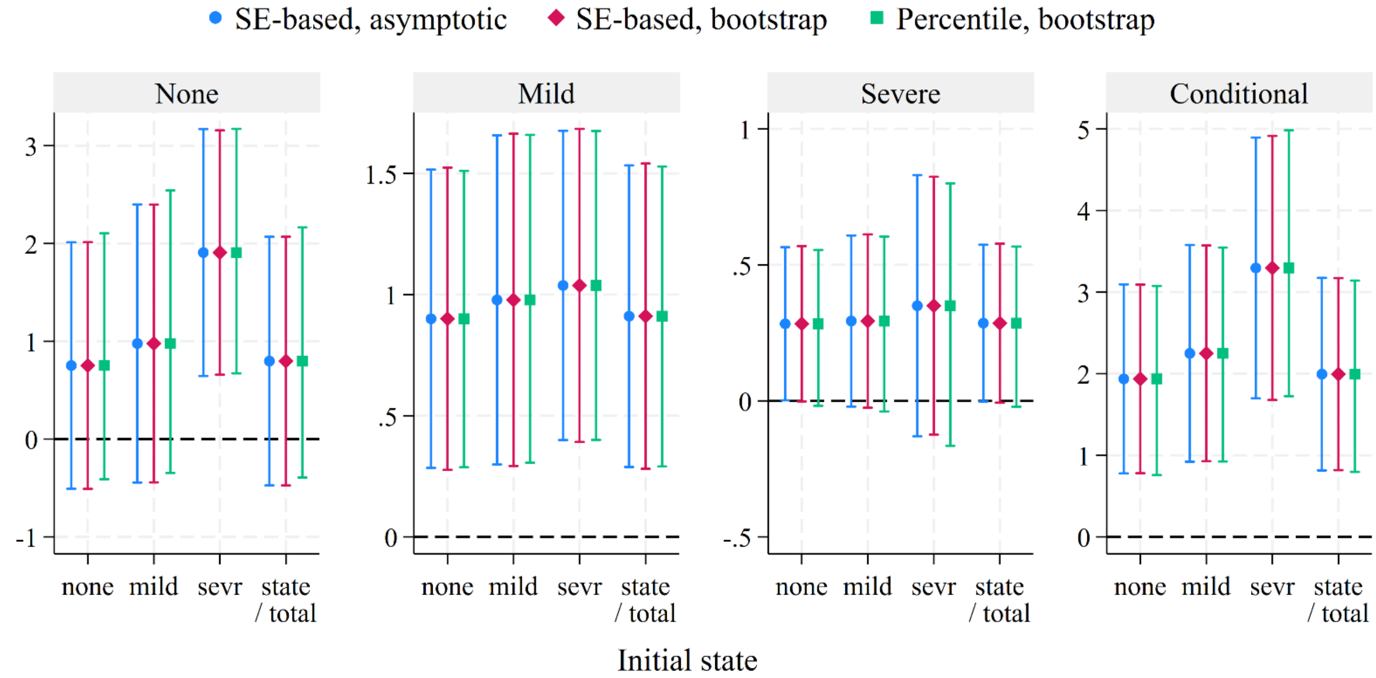
6.5 N-GROUP COMPARISONS

The formulas in sections 6.1-6.4 are for comparisons of two groups. They generalize to N-group comparisons in the obvious way.

6.6 COMPARISON TO BOOTSTRAP RESULTS

As one example for group comparisons, Figure 4 shows asymptotic CIs and their bootstrap counterparts for the difference between two sets of LEXP estimates using expectancies from the cognitive impairment example. A first LEXP result is based on predicted transition probabilities for women, and a second one for men. The statistical difference between the two is based on the joint covariance matrix covering both sets of estimates. The figure shows that discrepancies between the asymptotic and bootstrap CIs are generally very small.

Figure 4: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: Sex Difference (Women minus Men) in State and Overall Life Expectancies



Notes: Points depicted show the sex difference in state and overall life expectancies, calculated as the expectancies of women minus the expectancies of men. Otherwise notes from Figure 1 apply.

In a second illustration, we highlight a particular useful concept that the combination of LEXP and MAFN results holds: The division of LEXP estimates by lifetime risk (the latter is part of MAFN estimates). To see the practical value of this operation, consider the state expectancy of severe impairment. This is the expected life time in severe impairment for the average subject – and the "average" includes those subjects that never actually get severely impaired. Dividing this number by the lifetime risk of severe impairment one obtains the expected lifetime spent in severe impairment by those subjects who actually ever enter that state during their lifetimes, which undoubtedly is a number of interest. By building a joint covariance matrix of LEXP and MAFN estimates, one can immediately obtain asymptotic CIs for this division of LEXP results components by MAFN results components. Below, we perform this division for subjects that are assumed to be unimpaired at the base age, i.e., we divide the state expectancy of severe impairment, conditional on being healthy at the base age, by the lifetime risk numbers from section 5.3.

The transformation that we examine below, however, has a second aspect: We illustrate a more complex transformation that involves more than two sets of results. In principle, the number of results that can enter the joint covariance matrix is unlimited and only constrained by computation time. Here we divide the state expectancy of severe impairment by the lifetime risk of becoming severely impaired separately for women and men, and then calculate the difference between the resulting numbers. This involves a total of four results sets (LEXP and MAFN, separately for women and men). A joint covariance matrix for a larger set of results (12 results) is presented in section 8.

The upper section of Table 1 below lists point estimates and asymptotic CIs for all steps of this transformation. Healthy women and men at the base age have a severe state expectancy of 0.906 and 0.622, respectively, which yields a difference of 0.284, which is slightly under the 95% significance threshold. The lifetime risk of severe impairment of these subgroups is 0.316 and 0.246, respectively. When dividing these numbers separately for each group, we obtain impairment expectancies, conditional of becoming impaired, of 2.864 and 2.527 for women and men, respectively, the difference of which is 0.336. This difference is, contrary to the difference of the impairment expectancy that is not conditional on becoming impaired, significant at the 95% level.

The second section of Table 1 shows the same set of bootstrap point estimates and CIs. Here bootstrap point estimates are the averages across all replications, and CIs are based on the bootstrap standard error of the coefficients. The second section also shows the relative difference to corresponding asymptotic numbers, whose low values confirm, at a glance, a high degree of agreement between asymptotic and bootstrap numbers. Finally, the third section of Table 1 is similar to section two, except that point estimates are the median of coefficients over all bootstrap replications, and CIs are based on their 2.5th and 97.5th percentiles. The level of agreement with asymptotic numbers is very similar to that of section two of the table.

Table 1: Point Estimates, Asymptotic and Bootstrap 95% Confidence Intervals for Severe State Expectancy, Lifetime Risk for Severe Impairment, the Ratio of the Two, and the Difference of Ratios

	Women			**Men**			**Difference**		
	Point estimate	95% CI lower bound	95% CI upper bound	Point estimate	95% CI lower bound	95% CI upper bound	Point estimate	95% CI lower bound	95% CI upper bound
Asymptotic									
LEXP	0.906	[0.701	1.110]	0.622	[0.430	0.814]	0.284	[0.002	0.565]
LRSK	0.316	[0.274	0.359]	0.246	[0.204	0.289]	0.070	[0.012	0.128]
LEXP / LRSK	2.864	[2.491	3.236]	2.527	[2.102	2.952]	0.336	[-0.141	0.814]
Bootstrap, SE-based									
LEXP	0.887	[0.677	1.096]	0.608	[0.416	0.800]	0.279	[-0.007	0.564]
LRSK	0.313	[0.268	0.359]	0.243	[0.199	0.287]	0.070	[0.009	0.131]
LEXP / LRSK	2.822	[2.456	3.187]	2.486	[2.062	2.910]	0.336	[-0.141	0.812]
Relative difference to asymptotic results									
LEXP	0.010	[0.014	0.007]	0.009	[0.010	0.008]	0.004	[0.009	0.000]
LRSK	0.002	[0.004	0.000]	0.002	[0.004	0.001]	0.000	[0.003	0.003]
LEXP / LRSK	0.011	[0.010	0.012]	0.012	[0.013	0.011]	0.000	[0.000	0.001]

	Women			Men			Difference		
Bootstrap, percentile									
LEXP	0.888	[0.693	1.103]	0.604	[0.419	0.806]	0.276	[-0.018	0.555]
LRSK	0.313	[0.268	0.361]	0.244	[0.199	0.287]	0.070	[0.009	0.133]
LEXP / LRSK	2.824	[2.471	3.196]	2.485	[2.053	2.942]	0.340	[-0.141	0.782]
Relative difference to asymptotic results									
LEXP	0.009	[0.005	0.003]	0.011	[0.008	0.004]	0.006	[0.020	0.006]
LRSK	0.002	[0.005	0.001]	0.001	[0.004	0.002]	0.000	[0.004	0.004]
LEXP / LRSK	0.010	[0.006	0.010]	0.012	[0.016	0.002]	0.003	[0.000	0.018]

Notes: Comparison of asymptotic and bootstrap point estimates and 95% CIs for a set of four results (LEXP, separately for women and men, and MAFN, of which LRSK is a component result, separately for women and men). and linear and nonlinear transformation of them. The section labeled "Bootstrap, SE-based" holds point estimates based on the average of coefficients, calculated over all bootstrap replications, and 95% CIs are based on 1.96 times the bootstrap standard error of coefficients. The table section labeled "Bootstrap, percentile" has point estimates that are the medians of the bootstrapped coefficients, along with 95% percentile intervals. Note that the bootstrap sections of the table employ slightly different calculations than the figures in this document with respect to the point estimates, where point estimates that are depicted are asymptotic ones in all cases. Table rows labeled LEXP hold numbers for the severe impairment expectancy, conditional on no impairment at base age. Rows labeled LRSK hold the corresponding lifetime risk. Rows labeled LEXP / LRSK show the ratio of the two. The subsections that show relative difference numbers employ the relative difference formula $|a-b|/(|b|+1)$, where a and b are the values from the asymptotic and bootstrap sections, respectively.

7 Partial Age Ranges

A partial age range is defined by two ages z_b and z_e, called the beginning age and ending age, with $b, e \in [1, \ldots, \bar{a}]$ and $z_b < z_e$. We call a pair of age ranges disjunct if $e_1 \leq b_2$ (note the weak inequality), where the subscript indexes the age range.

7.1 Life Expectancy

Formulas for life expectancy apply unchanged. The partial age range is solely introduced via a redefinition of the rewards vector (15): The partial age range rewards vector $\boldsymbol{r}^{b,e}$ resets $n_a = 0$ if $z_a < z_b$ or if $z_a \geq z_e$, and is equal to $\boldsymbol{r}$ otherwise. We call an age partition a set of k disjunct age ranges for which it holds that $\sum_k \boldsymbol{r}^{b_k e_k} = \boldsymbol{r}$. Note that superscripts indicate age range information and not powers.

It shall be shown in the following that adding results of any two age ranges $\boldsymbol{r}^{b_1 e_1}$ and $\boldsymbol{r}^{b_2 e_2}$ is equivalent to specifying these age ranges in a single time rewards vector $\boldsymbol{r}^{b_1 e_1} + \boldsymbol{r}^{b_2 e_2}$. Without loss of generality, we show this for two adjacent age ranges for which $e_1 = b_2$, so that $\boldsymbol{r}^{b_1 e_1} + \boldsymbol{r}^{b_2 e_2} = \boldsymbol{r}^{b_1 e_2}$. By implication, adding the results (i.e., forming a simple additive linear combination) of all age ranges of an age partition is then equivalent to the result calculated in one step on the full age range, both in terms of point estimates and in terms of covariance matrices.

For the point estimates (18), the above statement can be immediately deduced:

$$
\begin{aligned}
\boldsymbol{E}_{b_1e_1} + \boldsymbol{E}_{b_2e_2} &= \boldsymbol{F}_1\left(\boldsymbol{r}^{b_1e_1\prime} \otimes \boldsymbol{I}_{\bar{s}}\right) + \boldsymbol{F}_1\left(\boldsymbol{r}^{b_2e_2\prime} \otimes \boldsymbol{I}_{\bar{s}}\right) \\
&= \boldsymbol{F}_1\left(\left(\boldsymbol{r}^{b_1e_1\prime} + \boldsymbol{r}^{b_2e_2\prime}\right) \otimes \boldsymbol{I}_{\bar{s}}\right) \\
&= \boldsymbol{F}_1\left(\boldsymbol{r}^{b_1e_2\prime} \otimes \boldsymbol{I}_{\bar{s}}\right) \\
&= \boldsymbol{E}_{b_1e_2}
\end{aligned} \tag{64}
$$

Noting that the two age range results are based on the same $\boldsymbol{V}^{F_1}$, recalling (58), and using the shorthand $\check{\boldsymbol{r}}_{+1}^{b_1e_1} = \boldsymbol{I}_{\bar{s}^2} \otimes \boldsymbol{r}_{+1}^{b_1e_1}$ and likewise for the second age range, we see that applying a simple additive linear combination to the joint covariance matrix of $\boldsymbol{E}_{b_1e_1}$ and $\boldsymbol{E}_{b_2e_2}$ yields

$$
\begin{aligned}
&[\boldsymbol{I}_{\bar{s}^2} \quad \boldsymbol{I}_{\bar{s}^2}]\mathrm{cov}(\mathrm{vec}[\boldsymbol{E}^{b_1e_1} \quad \boldsymbol{E}^{b_2e_2}]) \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \\ \boldsymbol{I}_{\bar{s}^2} \end{bmatrix} \\
&= [\boldsymbol{I}_{\bar{s}^2} \quad \boldsymbol{I}_{\bar{s}^2}] \begin{bmatrix} \check{\boldsymbol{r}}_{+1}^{b_1e_1} & \boldsymbol{0} \\ \boldsymbol{0} & \check{\boldsymbol{r}}_{+1}^{b_2e_2} \end{bmatrix} \left(\boldsymbol{1}_{2,2} \otimes \boldsymbol{V}^{F_1}\right) \begin{bmatrix} \check{\boldsymbol{r}}_{+1}^{b_1e_1} & \boldsymbol{0} \\ \boldsymbol{0} & \check{\boldsymbol{r}}_{+1}^{b_2e_2} \end{bmatrix}' \begin{bmatrix} \boldsymbol{I}_{\bar{s}^2} \\ \boldsymbol{I}_{\bar{s}^2} \end{bmatrix} \\
&= \left[\check{\boldsymbol{r}}_{+1}^{b_1e_1} \quad \check{\boldsymbol{r}}_{+1}^{b_2e_2}\right]\left(\boldsymbol{1}_{2,2} \otimes \boldsymbol{V}^{F_1}\right)\left[\check{\boldsymbol{r}}_{+1}^{b_1e_1} \quad \check{\boldsymbol{r}}_{+1}^{b_2e_2}\right]' \\
&= \check{\boldsymbol{r}}_{+1}^{b_1e_1}\boldsymbol{V}^{F_1}\check{\boldsymbol{r}}_{+1}^{b_1e_1\prime} + \check{\boldsymbol{r}}_{+1}^{b_2e_2}\boldsymbol{V}^{F_1}\check{\boldsymbol{r}}_{+1}^{b_1e_1\prime} + \check{\boldsymbol{r}}_{+1}^{b_1e_1}\boldsymbol{V}^{F_1}\check{\boldsymbol{r}}_{+1}^{b_2e_2\prime} + \check{\boldsymbol{r}}_{+1}^{b_2e_2}\boldsymbol{V}^{F_1}\check{\boldsymbol{r}}_{+1}^{b_2e_2\prime} \\
&= \left[\check{\boldsymbol{r}}_{+1}^{b_1e_1} + \check{\boldsymbol{r}}_{+1}^{b_2e_2}\right]\boldsymbol{V}^{F_1}\left[\check{\boldsymbol{r}}_{+1}^{b_1e_1} + \check{\boldsymbol{r}}_{+1}^{b_2e_2}\right]' \\
&= \check{\boldsymbol{r}}_{+1}^{b_1e_2}\boldsymbol{V}^{F_1}\check{\boldsymbol{r}}_{+1}^{b_1e_2\prime} \\
&= \mathrm{cov}(\mathrm{vec}\,\boldsymbol{E}^{b_1e_2})
\end{aligned} \tag{65}
$$

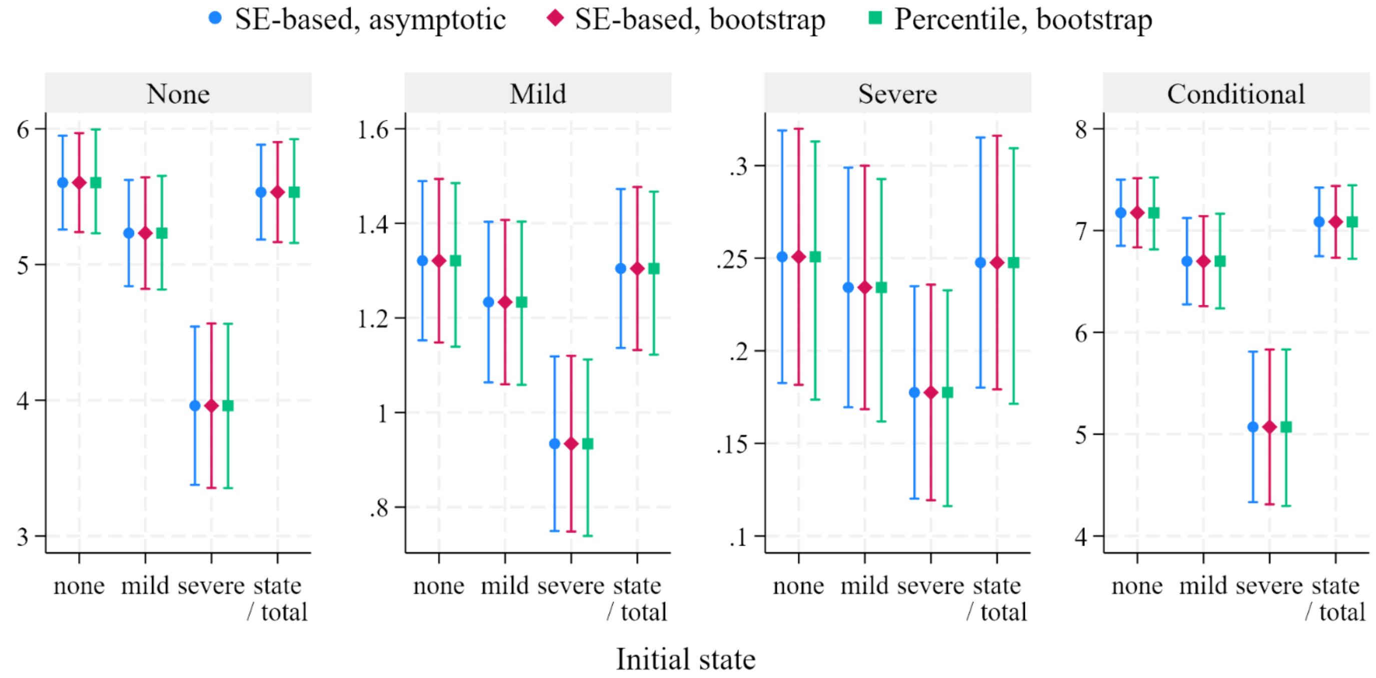
7.2 COMBINING PARTIAL AGE RANGE RESULTS

Partial age range results can be combined with other results in the usual way. The formulas of section 5.3 apply unchanged, with $\boldsymbol{r}^{b,e}$ taking the place of $\boldsymbol{r}$ whenever appropriate.

7.3 COMPARISON TO BOOTSTRAP RESULTS

Similar to Figure 1, Figure 5 shows state and overall life expectancies from the cognitive impairment example. The major difference is that Figure 5 shows numbers for the partial age range of 70-80. The agreement of asymptotic and bootstrap CIs is again very high.

Figure 5: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: State and Overall Life Expectancies, Partial Age Range 70-80, Women



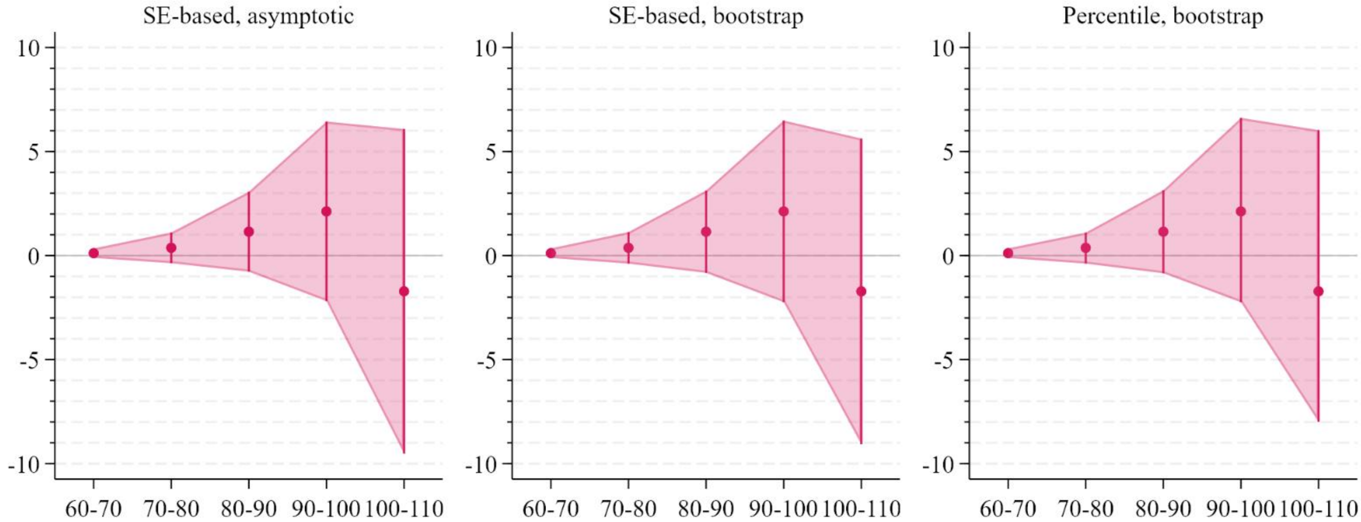
Notes: Points depicted show state and overall life expectancies for women and the partial age range 70-80. Otherwise notes from Figure 1 apply.

8 COMPLEX NONLINEAR COMBINATIONS AND JOINT HYPOTHESIS TESTS

This section draws upon results from sections 4, 6, and 7 and illustrates the possibility of asymptotic calculations of CIs for complex linear and nonlinear combinations based on many sets of results. The prerequisite for this is the calculation of a joint covariance matrix that covers all sets of results, as developed earlier.

We again use the cognitive impairment example. The starting point are state expectancy predictions for 10-year partial age ranges, separately for women and men. The six partial age ranges form a partition of the full age range 50-110. We then transform, separately for each partial age range and for women and men, the severe impairment expectancy into the percentage of lifetime in severe impairment by dividing by the total life expectancy. Next, separately for women and men, we calculate the percentage point increase in severe impairment expectancy by age decade. Finally, for each age range, we deduct the result for women from that for men. This procedure answers the following question: What are the age decades during which women's share of life spent in severe impairment increases particularly strongly in comparison to men? While this example may be seen as a little contrived, it illustrates the possibilities that the calculation of joint covariance matrices across any type and number of results holds. Figure 6 compares the results based on analytical calculations (subgraph on the left) to bootstrap results (middle and right subgraphs). The only slightly visible differences appear at the very highest age range (100-110), where data scarcity leads to imprecision of estimation.

Figure 6: Comparison of 95% Confidence Intervals, Asymptotic v. Bootstrap Results: Complex Linear and Nonlinear Combinations Using Partial Age Ranges



Notes: Each subgraphs shows sex differences (women minus men) for the percentage point increase, by age decade, in the fraction of time spent in severe impairment. Age decades are depicted on the horizontal axis. They are to be read as intervals [a, b), i.e., with a strong inequality on the right boundary.

The resulting CIs can be used to assess significance of individual transformations (individual age ranges). It is also possible to use the joint asymptotic covariance matrix for a joint test of several hypotheses. For example, one can ask whether all point estimates in (the left subgraph of) Figure 6 are zero. Visual inspection would suggest a positive answer. A corresponding asymptotic Wald test yields a $\mathcal{X}^2(5)$ statistic of 10.3 with associated p-value of 0.067, which would reject the hypothesis at the 10% level. This can be compared to the SE-based bootstrap results using the covariance matrix of coefficients calculated over the coefficient estimates of all bootstrap replications. A Wald test based on this covariance matrix yields a $\mathcal{X}^2(5)$ statistic of 8.08 with associated p-value of 0.152. This is an instance where asymptotic and bootstrap results diverge and lead to different conclusions. The source of this difference remains to be investigated.

9 REFERENCES

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods* (1st ed.). MIT Press.

Dudel, C. (2021). Expanding the Markov Chain Toolbox: Distributions of Occupation Times and Waiting Times. *Sociological Methods & Research*, *50*(1), 401–428. https://doi.org/10.1177/0049124118782541

Dudel, C., & Schneider, D. C. (2021). How bad could it be? Worst-case bounds on bias in multistate models due to unobserved transitions. *Sociological Methods & Research*, *52*(4), 1816–1837. https://doi.org/10.1177/0049124121995540

Hunter, C. M., & Caswell, H. (2005). The use of the vec-permutation matrix in spatial matrix population models. *Ecological Modelling*, *188*(1), 15–21. https://doi.org/10.1016/j.ecolmodel.2005.05.002

Lièvre, A., Brouard, N., & Heathcote, C. (2003). The Estimation of Health Expectancies from Cross-Longitudinal Surveys. *Mathematical Population Studies*, *10*(4), 211–248. https://doi.org/10.1080/713644739

Lütkepohl, H. (1996). *Handbook of Matrices*. Wiley.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Lynch, S. M., & Brown, J. S. (2005). A New Approach to Estimating Life Tables with Covariates and Constructing Interval Estimates of Life Table Quantities. *Sociological Methodology*, *35*(1), 189–237. https://doi.org/10.1111/j.0081-1750.2006.00168.x

Millimet, D. L., Nieswiadomy, M., Ryu, H., & Slottje, D. (2003). Estimating worklife expectancy: An econometric approach. *Journal of Econometrics*, *113*(1), 83–113. https://doi.org/10.1016/S0304-4076(02)00168-9

Roth, G., & Caswell, H. (2018). Occupancy time in sets of states for demographic models. *Theoretical Population Biology*, *120*, 62–77. https://doi.org/10.1016/j.tpb.2017.12.007

Schneider, D. C. (2023). *dtms: A Stata package to estimate discrete-time multistate models.* (0.3.1) [Stata]. https://user.demogr.mpg.de/schneider/

Schneider, D. C., Myrskylä, M., & van Raalte, A. (2023). Flexible transition timing in discrete-time multistate life tables using Markov chains with rewards. *Population Studies*. https://doi.org/10.1080/00324728.2023.2176535

Wooldridge, J. M. (2002). *Econometric analysis of cross-section and panel data* (1st ed.). MIT Press.

10 APPENDIX

10.1 NAÏVE APPROACH OF CALCULATING $\boldsymbol{V^F}$

Equation (24) in section 4.2 gave an expression for the covariance matrix of the full fundamental matrix. This section derives this expression.

Since $\boldsymbol{F} = \boldsymbol{g}(\boldsymbol{U})$, see (16), the delta method tells us that, asymptotically,

$$\boldsymbol{V}^F = \boldsymbol{G}^F \boldsymbol{V}^U \boldsymbol{G}^{F'}$$

with

$$\boldsymbol{G}^F = \frac{\partial \operatorname{vec} \boldsymbol{F}}{\partial \operatorname{vec}(\boldsymbol{U})'} = \frac{\partial \operatorname{vec}\left((\boldsymbol{I}-\boldsymbol{U})^{-\mathbf{1}}\right)}{\partial \operatorname{vec}(\boldsymbol{U})'} = \frac{\partial \operatorname{vec}\left((\boldsymbol{I}-\boldsymbol{U})^{-\mathbf{1}}\right)}{\partial \operatorname{vec}((\boldsymbol{I}-\boldsymbol{U}))'} * \frac{\partial \operatorname{vec}\left((\boldsymbol{I}-\boldsymbol{U})\right)}{\partial \operatorname{vec}(\boldsymbol{U})'}$$

Noting that that, for any invertible matrix $\boldsymbol{Z}$,

$$\frac{\partial \operatorname{vec}(\boldsymbol{Z}^{-\mathbf{1}})}{\partial \operatorname{vec}(\boldsymbol{Z})'} = -\boldsymbol{Z}^{-1'} \otimes \boldsymbol{Z}^{-1}$$

(see, e.g., Lütkepohl, 2005, p.668), the first term resolves to

$$\frac{\partial \operatorname{vec}\left((\boldsymbol{I}-\boldsymbol{U})^{-\mathbf{1}}\right)}{\partial \operatorname{vec}((\boldsymbol{I}-\boldsymbol{U}))'} = -(\boldsymbol{I}-\boldsymbol{U})^{-1'} \otimes (\boldsymbol{I}-\boldsymbol{U})^{-1} = -\boldsymbol{F}' \otimes \boldsymbol{F}$$

Since the second term is simply

$$\frac{\partial \operatorname{vec}\left((\boldsymbol{I}-\boldsymbol{U})\right)}{\partial \operatorname{vec}(\boldsymbol{U})'} = -\boldsymbol{I}_{(\overline{s}\overline{\boldsymbol{a}}_{-1})^2}$$

we get

$$\boldsymbol{G}^F = \boldsymbol{F}' \otimes \boldsymbol{F}$$

and

$$\boldsymbol{V}^F = (\boldsymbol{F}' \otimes \boldsymbol{F})\boldsymbol{V}^U(\boldsymbol{F}' \otimes \boldsymbol{F})'$$