ARC Centre of Excellence in Population Ageing Research

Working Paper 2023/01

Dynamics of Health Expectancy: An Introduction to Multiple Multistate Method (MMM)

Tianyu Shen, Collin Payne and Maria Jahromi

Dynamics of Health Expectancy:

An introduction to Multiple Multistate Method (MMM)

Tianyu Shen[1], *Collin F. Payne*[1,2] *and Maria Jahromi*[3]

3 January 2023

Abstract

Many studies have compared individual measures of health expectancy across older populations by time-invariant variables. However, very few have included time-varying variables when calculating health expectancy. Since events in the life course are likely to be changing over time in related ways, it is valuable to incorporate time-varying socioeconomic factors. This paper proposes a Multiple Multistate Method (MMM) that situates the multistate model within the broader family of Vector Autoregression (VAR) models. When estimating multistate models with sample survey data, sparseness in the transition matrices often makes such models unfeasible should two or more time-varying variables be built into the state spaces. This approach allows for the estimation of more complex state spaces (including the modeling of time-varying covariates) by reducing less important interactions in the model. We then demonstrate the MMM in two empirical applications, showing the flexibility of the approach to explore health expectancies with complex state spaces.

Keywords: multistate model, discrete-time Markov processes, microsimulation, health expectancy, VAR model

[1] School of Demography, Australian National University
[2] Harvard Center for Population and Development Studies, Harvard T.H. Chan School of Public Health
[3] Centre for Social Research and Methods, Australian National University

Correspondence to: tianyu.shen@anu.edu.au

Introduction

In recent years, a considerable body of research has developed using multistate models to explore health expectancies based on data from longitudinal sample surveys. Several main analytical approaches have been developed to estimate multistate life table quantities from longitudinal data, including the Stochastic Population Analysis for Complex Events (SPACE) program (Cai et al. 2010), the Interpolated Markov Chain Method (IMaCh) (Lièvre et al. 2003), and the Gibbs Sampler for Multistate Life Tables Software (GSMLT) (Lynch and Brown 2005). However, a shared challenge that these models face is their inability to handle large, complex state spaces—a shortcoming that is mostly due to the relatively small sample sizes available from longitudinal sample survey data. In multistate models, including a more refined categorization of health or having more than one time-varying variable leads to a rapid growth in the state space to be estimated. As this state space increases, "the number of transition schedules to be estimated increases multiplicatively" (Saito et al. 2014:216). This scaling issue leads to issues of sparsity, as observed transitions become rare and age-patterns difficult to estimate.

Due to these methodological challenges, existing studies have mostly computed health expectancy or other multistate life expectancies assuming an individual's socio-demographic factors remain constant over time. The literature focuses heavily on differences across time-invariant factors such as sex and education. A few studies have explored time-varying variables such as urban/rural residence (Liu et al. 2019) and marital status (Martikainen 2014) by assuming these variables remain unchanged in later life. A small body of recent studies have attempted to include time-varying variables by including them in the state space, an approach that we hereafter call the "complex multistate model". Jia and Lubetkin (2020) combined marital and disability status into a complex multistate model with two disability states and five different states of marital status. The resulting state space is extremely large, and many

tianyu.shen@anu.edu.au

transition probabilities need to be estimated. Estimating accurate transition probabilities for such a large, complex state space requires a massive amount of data, and Jia and Lubetkin (2020) use data from Medicare Health Outcomes Survey comprising over 160,000 respondents. Huang et al (2021) computed the older Chinese health expectancy of physically active and free of cognitive impairment in a complex multistate model. Yet, their state space omitted some of the possible states without further explanation. Another working paper by *Authors* (2022) also developed a multidimensional extension to prior work on health expectancy by simultaneously modelling changes in morbidity and disability across a set of cohorts in the US Health and Retirement Survey (HRS). However, they had to use a very simplified five-state state space (using binary measures of any vs. no morbidities and any vs. no disability) to estimate these quantities, even with the substantial sample size of the HRS.

An active strain of research on multistate methods has sought to overcome some of the limitations associated with estimating health expectancies in complex state spaces. A recent paper by Lynch and Zang (2022) uses a Bayesian approach to account for issues of data sparsity when estimating quantities in the complex multistate model. Other studies have also sought to incorporate time-varying variables using methods other than the traditional multistate model. Chiu (2019) computed the disability-free life expectancy by living arrangements in the US, claiming that living arrangement is treated as a time-varying covariate in the model. However, the description of the method is unclear about how this time-variant covariate was operationalized. Only one method --- the simultaneous equation system used by Yang and Hall (2007) --- appears to address this issue. This method estimates health expectancy with several time-variant covariates (i.e., BMI, medical events and chronic diseases) within a system of equations. Yet, little existing research has applied this method, partly due to its statistical complexity. Thus, the aim of this paper is to develop a simple and generalizable method to allow increased complexity in multistate state spaces when incorporating multiple time-varying

tianyu.shen@anu.edu.au

variables, such as multiple dimensions of health, or interactions between health and socioeconomic variables, when estimating health expectancies or other multistate life expectancies.

Conceptually speaking, in the multistate model, when an individual moves to a different state that individual assumes a new set of transition probabilities. A similar idea can be found in the complex multistate model, where a change in one of the time-varying impacts both its own transition probabilities and other time-varying variables' transition probabilities. For example, an individual becoming obese shifts not only their probability of whether they will be obese in the future but also their probability of developing diabetes.

In this paper, we introduce a formulation of the complex multistate model with more than one time-varying variable (e.g., *Authors* 2022; Jia and Lubetkin 2020) as a recursive vector autoregression (VAR) model. We call this new representation of complex multistate model the multiple multistate method (MMM). The concept of this modelling framework shares many similarities with the vector autoregressive (VAR) model popular in econometric time series studies, which is used to capture the relationship between multiple endogenous variables as they change over time. VAR models have not been used in in the context of modelling health expectancy before, although they have been applied in actuarial studies to forecast mortality (e.g., Chang and Shi 2021; Guibert, Lopez and Piette, 2019; Li and Lu, 2017; Li and Shi, 2021). In the methods section, we describe how the MMM can exactly replicate a complex multistate model and discuss how the flexibility of the MMM approach can reduce estimation difficulties by removing less important interactions when estimating complex state spaces.

To better illustrate our method, we present two examples. The first example replicates a working paper (*Authors* 2022). Using a five-state multistate model, they estimated the health expectancy by both morbidity and disability of the four successive US birth cohorts, born from

tianyu.shen@anu.edu.au

1914-1923 to 1944-1953. To demonstrate the method, we adopt the same data and compare results between the complex multistate model and the MMM. In the second example, we explore a similar research question to Jia and Lubetkin (2020). Instead of looking at marital status and ADL disability, we select another commonly used health indicator --- self-rated health (e.g., Crimmins 2004; Payne 2022). For decades, many studies have discussed the association between marital status and health. Most of them suggest a positive or protective effect of marriage on health and survival (Goldman et al. 1995; Rendall et al 2011, Verbrugge 1979). Others also found negative effects of widowhood or divorce (Korinek et al. 2011; Verbrugge 1979). Yet only very few studies examined the impact through the lens of multistate life expectancy until Jia and Lubetkin (2020). Thus, this example may provide dynamic insight into how marital status and health status interact as individuals age.

Method

Complex Multistate Model

To explore the interaction between two time-varying variables with, for example, two categories each, the traditional complex multistate method would combine the two variables to form five distinct states with one absorbing state for dead. The first time-varying Variable *G* has two categories, g_1 and g_2, and Variable *H* has h_1 and h_2. There are different approaches to estimate transition probabilities or rates for the expectancy (e.g., Allison 1982; Dudel 2021; Lynch and Brown 2005); the method in this paper is built on logistic regression, one of the most widely used discrete-time methods (Allison 1982; Cai et al. 2010). The state space is shown in Figure 1 and the corresponding Table 1 present the matrix of transition probabilities within the state space, where the row names represent the current state, and the column names the state in time *t+1*. Each row sums up to one.

tianyu.shen@anu.edu.au

Using multinomial regression, we can estimate the transition probabilities in equation (1),

$$\text{logit}(state_{t+1} = s|\mathbf{x}) = \ln\left[\frac{\Pr(state_{t+1} = s|\mathbf{x})}{\Pr(state_{t+1} = S|\mathbf{x})}\right] = \alpha_s + \beta_{1,s} \cdot state_t + \beta_{2,s} \cdot age_t + \beta_{i,s} \cdot \mathbf{X}_i, \quad s = \{g_1h_2, g_2h_1, g_2h_2, Dead\}; S = \{g_1h_1\} \quad (1)$$

where $\mathbf{X}_i$ may include other fixed covariates, such as sex or education. Multinomial logistic regression estimates a series of binary logistic models comparing each state, *s*, to the baseline state, *S* (set to g_1h_1), given all the covariates on the right-hand side (Agresti 2007; Fullerton and Xu 2018). This regression model assumes, as in prior literature (e.g., Cai and Lubitz 2007; Laditka and Wolf 1998; Lim et al. 2019), that the transitions follow a discrete-time Markov chain, and the transition probabilities are time-homogeneous and age-specific. A set of coefficients is estimated for each comparison to the baseline. The predicted probability of being in state, *s*, in time *t+1* given the current state and other covariates can be expressed in equation (2),

$$\Pr(state_{t+1} = s|state_t = c, age_t, \mathbf{X}) = \frac{e^{\alpha_s + \beta_{1,s} \cdot state_t + \beta_{2,s} \cdot age_t + \beta_{i,s} \cdot \mathbf{X}_i}}{\sum_h e^{\alpha_h + \beta_{1,h} \cdot state_t + \beta_{2,s} \cdot age_t + \beta_{i,h} \cdot \mathbf{X}_i}}, \quad s = \{g_1h_1, g_1h_2, g_2h_1, g_2h_2, Dead\}; c = \{g_1h_1, g_1h_2, g_2h_1, g_2h_2\}, \quad (2)$$

where *h* includes all five possible states, and **X** includes all the covariates other than $state_t$ and age_t in equation (1).

After obtaining this transition matrix, we can use microsimulation to calculate the life/health expectancy. To do this, we generate 100,000 individuals based on the baseline characteristics of the start age of the 10-year age group. The probabilities in each row of Table 1 are mapped into subsets in the interval of 0 to 1 based on the size of each probability. For example, the first row would be turned into five subsets: $[0, \mu_1)$, $[\mu_1, \mu_1 + \mu_2)$, $[\mu_1 + \mu_2, \mu_1 + \mu_2 + \mu_3)$, $[\mu_1 + \mu_2 + \mu_3, \mu_1 + \mu_2 + \mu_3 + \mu_4)$ and $[\mu_1 + \mu_2 + \mu_3 + \mu_4, 1]$. Then a random

tianyu.shen@anu.edu.au

number, *X*, is drawn from the uniform distribution, $X \sim U(0,1)$. The next state of the individual with certain current state and other characteristics is assigned to whichever subset this random number falls into (Laditka and Wolf 1998). Life/health expectancy can also be calculated using multistate life table method, with the same transition probabilities and baseline characteristics. However, the synthetic cohort from microsimulation can provide much richer information on individuals' life courses, beyond simple estimates of aggregated life expectancy.

(Table 1 & Figure 1 about here)

With the basis of the traditional multistate model explained, we can introduce the Multiple Multistate Method (MMM) and highlight its distinct features. As mentioned in the Introduction, one of the main drawbacks of using the traditional multistate model is its inability to handle large state spaces. In the case of two categories in each dimension of health and five states in total, the traditional multistate model is still manageable. However, when the number of categories or the number of time-varying variables increases, the observed transitions often become too sparse to reliably estimate. To alleviate this problem, MMM proposes to model different time-varying variables separately in multiple logistic regressions.

Multiple Multistate Method (MMM)

The concept of MMM shares many similarities with the vector autoregressive (VAR) model, which is commonly used in macroeconomic and financial modelling to capture the relationship between multiple endogenous variables as they change over time. The advantage of VAR models over other regression models such as ordinary least squares is that they do not require variables to be exogenous (Enders 2004), so these models are well suited for modelling interrelated variables such as macroeconomic variables. In our context, this means that a VAR

tianyu.shen@anu.edu.au

model can jointly estimate a system of equations where, for example, disability can be modelled as a function of morbidity, and morbidity can be modelled as a function of disability. A VAR model usually takes one of three forms: *reduced-form* VAR, *recursive* VAR and *structural* VAR (SVAR) (Stock and Watson 2001). In a *reduced-form* VAR, each variable is modelled as a function of its own past and the past values of the other variables (i.e., the lags of the variables), but the model does not capture the contemporaneous effects (Enders 2004). On the other hand, *recursive* and *structural* VAR models include the lags of the variables similar to a reduced-form VAR, but in addition, they also allow the outcome variables in each equation to depend on the contemporaneous values of the other variables. The *structural* VAR differs from the *recursive* VAR in that it incorporates identifying assumptions derived from empirical theory and therefore allows for causal inference (Stock and Watson 2020). The *reduced-form* and *recursive* VAR models are more relevant for the purpose of this paper, as our focus is not on causal inference. The advantage of a *reduced-form* VAR is in its simplicity, which is also its disadvantage. Since it does not include the contemporaneous variables, the short-run concurrent relationship or shock in the system would be ignored. Thus, whether *reduced-form* VAR is a good model depends on correlation of the change in one variable on the other at the same time and the interval of time unit.

There are applications using VAR on panel data (i.e., a Panel VAR model) even though the observation period may be shorter than 10 years (Canova and Ciccarelli, 2013; Holtz-Eakin et al. 1988; Kim and Lee, 2008). Most demographic models using longitudinal data to estimate multistate life tables would pool all individuals over time for the estimation of transition probabilities (e.g., Cai and Lubitz 2007; Payne 2022; Yong and Saito 2012), which provides more units of observation and lower complexity than the model with individual effects. VAR models are usually estimated with continuous time-series variables such as macroeconomic and financial data, where each equation in the VAR system is estimated via Ordinary Least

tianyu.shen@anu.edu.au

Squares regressions. However, in our context, the variables are binary, and we therefore estimate the equations via logit regressions. Such "logistic VAR models" have previously been applied in empirical work (e.g., Epskam 2013; Huang et al. 2020).

In economics, the lag length for the variables in each equation is typically estimated via F-tests or information criteria (such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC)) and will also depend on the frequency of the time-series variables (Stock and Watson 2020). For example, the number of lags is typically small with annual data, with around one or two lags (Wooldridge 2020). Using one lag is consistent with the common multistate assumption of a first-order Markov Chain, and annual or biannual survey data collection. With the Markov assumption, the current state depends only on the previous state, which can be regarded as a univariate autoregression with lag one, VAR(1). When the state spaces are the combinations of two variables (a complex multistate), it is possible to turn it into bivariate autoregressions with lag one maintaining the Markov assumption. As aforementioned, there are various types of VAR models, and the traditional multistate model is close to a *recursive* VAR because the concurrent relationship is estimated as one of the probabilities. In the following paragraphs, we will first explain the comparability between the complex multistate model and the bivariate *recursive* VAR(1), and then discuss and examine the potential alternative, *reduced-form* VAR(1) in the Application section.

The first problem to separately estimate the time-varying variables is to deal with the transition to mortality. A person can possess multiple time-varying characteristics at the same time, but there is only one dead state in Figure 1. It is possible to estimate both models with death and the transition probabilities to death for the same group of individuals should be equivalent because the number of transitions to death are the same in either time-varying characteristic. The probabilities to death are theoretically equivalent, so death should only be modelled in one of the time-varying variables. As a result, there will be two (or more) ways to

tianyu.shen@anu.edu.au

build the model depending on where the death is modelled. Nevertheless, these models are comparable in mathematics and results. To demonstrate this, we describe one type of model in the main text and detail the other one in the Appendix 1. Type a, in Figure 2, is where variable *G* can transition to death and variable *H* does not, whereas Type b, in Appendix Figure A1, flips this and only variable *H* can transition to death. Table 2 and A1 are the corresponding transition matrices for each type of model. Both tables have two panels. Panel a represents the transitions in the upper model and panel b the lower one. Each row should also sum to one.

The row names in Table 2 and A1 represent the current state of the time-varying variable that is modelled, and state at time t (and $t+1$) of the other time-varying variable is in subscript. To estimate these probabilities, we could model separately for each subscript or by adding interaction terms with all covariates. In this study, we choose to model Type a separately in equation (3) and (4) for simplicity (Type b can be found in Appendix 1),

$$\text{logit}\left(G_{t+1,h_t}\right) = \alpha_{h_t} + \beta_{1,h_t} G_{t,h_t} + \beta_{2,dis_t} age_{t,h_t} + \beta_{i,h_t} \mathbf{X}_{i,h_t}, \qquad h_t = \{h_1, h_2\} \qquad (3)$$

$$\text{logit}\left(H_{t+1,g_t,g_{t+1}}\right) = \alpha_{g_t,g_{t+1}} + \beta_{1,g_t,g_{t+1}} H_{t,g_t,g_{t+1}} + \beta_{2,g_t,g_{t+1}} age_{t,g_t,g_{t+1}} + \beta_{i,g_t,g_{t+1}} \mathbf{X}_{i,g_t,g_{t+1}}, \qquad g_t, g_{t+1} = \{g_1, g_2\} \qquad (4)$$

Thus, there are in total six regressions to estimate, two for G_{t+1} given h_t and four for H_{t+1} given g_t and g_{t+1}. Additionally, $state_t = g_1 h_1$ are the same observations as $G_{t,h_t=h_1} = g_1$, and so on.

(Table 2 & Figure 2 about here)

tianyu.shen@anu.edu.au

If the MMM can exactly replicate the complex multistate model, transition probabilities in Table 2 should have equivalent relation with those in Table 1. Pr $(G_{t+1,h_t=h_1} = g_1 | G_{t,h_t=h_1} = g_1, \mathbf{X})$, or $\lambda_{h_1.1}$, is the joint probability of $\Pr(state_{t+1} = g_1h_1 | state_t = g_1h_1, \mathbf{X})$, or μ_1, and $\Pr(state_{t+1} = g_1h_2 | state_t = g_1h_1, \mathbf{X})$, or μ_2. Since μ_1 and μ_2 are mutually exclusive, the joint probability is $\mu_1 + \mu_2$. Similarly, Pr $(G_{t+1,h_1} = g_2 | G_{t,h_1} = g_1, \mathbf{X})$, or $\lambda_{h_1.2}$, is the joint probability of μ_3 and μ_4 (i.e., $\mu_3 + \mu_4$). Pr $(G_{t+1,h_1} = \text{Dead} | G_{t,h_1} = g_1, \mathbf{X})$, or $\lambda_{h_1.3}$ is equivalent to $\Pr(state_{t+1} = \text{Dead} | state_t = g_1h_1, \mathbf{X})$, or μ_5.

(Table 3 about here)

As for panel b in Table 2, the probability of the transition in Variable *H* is conditional on the transition in Variable *G*, both time t and $t+1$. Thus, Pr $(H_{t+1,g_t=g_1,g_{t+1}=g_1} = h_1 | H_{t,g_t=g_1,g_{t+1}=g_1} = h_1, \mathbf{X})$, or $\lambda_{g_1.g_1.1}$, is equivalent to the $\frac{\mu_1}{\mu_1+\mu_2}$, and $\lambda_{g_1.g_1.2}$ is equivalent to $\frac{\mu_2}{\mu_1+\mu_2}$. Table 3 presents the equivalence of all probabilities between Tables 1 and 2. Noted that we use "equivalent" throughout these paragraphs because all these probabilities are predicted estimates instead of the actual probabilities and they are likely to be slightly different between models.

The dashed arrow that connects the two models can also be done in two ways like the complex multistate model: life table method or microsimulation. In this paper, we only focus on the microsimulation method, but the underlying calculation is the same for life table method. To get the life/health expectancies, we need to run two independent microsimulations for each type of model as there are two sets of transition probabilities. For an individual starting from

tianyu.shen@anu.edu.au

g_1 and h_1, we first draw a random number from the uniform distribution, $U(0,1)$, to determine the next states of Variable *G* according to which subset, in subsets $[0, \lambda_{h_1.1})$, $[\lambda_{h_1.1}, \lambda_{h_1.1} + \lambda_{h_1.2})$ and $[\lambda_{h_1.1} + \lambda_{h_1.2}, 1]$, the random number falls into. If the number falls in the first subset, this individual will become g_1. Thus, another random number is drawn, from the same uniform distribution, to identify the next state of Variable *H*, in subsets $[0, \gamma_{g_1.g_1.1})$ and $[\gamma_{g_1.g_1.1}, 1]$, this random number falls for the next. In such, we obtain the next state of Variable *G* and *H* for this individual. Alternatively, if the number falls in the third subset, then the second random number is not needed since this individual remains dead in all the future states.

So far, this model of MMM can exactly replicate the complex multistate model, but it does not solve the problem of the sparse transitions in certain category, especially when the sample size of the survey is small. The real benefit of MMM disentangling the complex state space by dividing one regression in equation (1) to multiple regressions in equations (3) & (4) is to enable higher level of flexibility. This makes estimation of more complex models possible by allowing the researcher to reduce insignificant interactions and manipulate the relationship between time-varying variables according to their research questions and theoretical frameworks.

Applications

In this section, we first demonstrate the multiple multistate method (MMM) that can exactly replicate a five-state multistate model focusing on two dimensions of health as in *Authors* (2022). We then demonstrate how the *reduced-form* VAR model can be used to estimate multistate life table quantities in complex state spaces, reducing estimation difficulties through removing less important interactions. The results from these methods are compared to examine the differences and understand potential limitations. We then present a further example showing how our method can be applied using time-varying variables other than health. This

tianyu.shen@anu.edu.au

example uses the MMM to explore healthy life expectancy while accounting for changes in marital status. All analyses are conducted in R software (R Core Team 2022).

Data

Data are from the US Health and Retirement Survey (HRS) (Health and Retirement Study 2021), a bi-annual national longitudinal survey (Sonnega et al. 2014) for both application examples. In example one, our analyses use data from 1998 to 2018 of the HRS to estimate cohort partial health expectancy with disability and morbidity across birth cohorts. Disability and morbidity are defined the same way as in *Authors* (2022). Disability is classified into two categories: "Disability-free" (DF) and "(Activities of Daily Living) ADL disabled" (D). Individuals are classified as "Morbid" (M) if they are ever diagnosed with any of the five chronic diseases including cancer, diabetes, heart disease, lung disease and stroke, and "Morbidity-free" (MF) otherwise.

In the second example, we use data from the 2008 to 2018 waves of the HRS to estimate remaining healthy life expectancy by sex and marital status for those aged 55 plus. Marital status is divided into three categories: "married/partnered", "divorced/ separated" and "widowed/widower". Individuals who never married are excluded from the analyses as they are a very small population and are unlikely to change marital status over time in older cohorts. Health is defined by self-rated health, where individuals who responded "Excellent" or "Very good" are reclassified as "very good", "Good" are "fair", and "Fair" or "Poor" are "poor".

Example 1: two dimensions of health

Figure 3 describes the complex multistate model and state space of *Authors* (2022), replacing the variables *G* & *H* in the example above to Morbidity and Disability. Note that the state space is slightly constrained as compared to Figure 1, as transitions from morbid to morbidity-free

are not allowed under the definition of morbidity as ever diagnosed. In this complex multistate model, the transition probabilities are estimated using a single multinomial logistic regression shown in equation (5).

$$\text{logit}(state_{t+1} = s|\mathbf{x}) = \ln\left[\frac{\Pr(state_{t+1} = s|\mathbf{x})}{\Pr(state_{t+1} = S|\mathbf{x})}\right] = \alpha_s + \beta_{1,s} \cdot state_t + \beta_{2,s} \cdot age_t + \beta_{i,s} \cdot \mathbf{X}_i, \quad s = \{\text{MF-D, M-DF, M-D, Dead}\};\ S = \{\text{MF-DF}\} \tag{5}$$

where $\mathbf{X}_i$ includes terms for age-squared, sex, birth cohorts and interactions between age, sex and birth cohorts.

(Figure 3 about here)

We omit the details of how to exactly replicate this complex multistate model using MMM with recursive VAR(1) in the main text as the procedure is highly similar to what is described in the Method section (for the interested reader, the detailed procedure can be found in Appendix 2). Instead, in this section, we focus on describing and comparing one of the potential alternatives that utilizes features of MMM to reduce complexity in the state space. Instead of modelling morbidity by disability group and disability by morbidity group like equations (3) and (4), we add them into the independent variables. To retain some of the non-linear relationship between age, morbidity and disability at time *t*, extra interaction terms are included. However, the other variables no longer have the interactions with health states. The MMM would be estimated by equations (6) and (7).

$$\text{logit}(morbidity_{t+1}) = \alpha_m + \beta_{m1} morbidity_t + \beta_{m2} disability_t + \beta_{m3} age_t + \beta_{m4} morbidity_t * age_t + \beta_{m5} disability_t * age_t + \beta_{mi} \mathbf{X}_i, \tag{6}$$

tianyu.shen@anu.edu.au

$$\text{logit}(disability_{t+1}) = \alpha_d + \beta_{d1} morbidity_t + \beta_{d2} disability_t + \beta_{d3} age_t + \beta_{d4} morbidity_t * age_t + \beta_{d5} disability_t * age_t + \beta_{di} \mathbf{X}_i, \quad (7)$$

where $\mathbf{X}_i$ similarly includes terms for age-squared, sex, birth cohorts and interactions between age, sex and birth cohorts, as well as interaction between sex, morbidity and disability.

As discussed above, only one of the equations allows for transitions to mortality, although the choice of which state individuals can die from does not matter for the final estimates (as shown in Appendix 1). Figure 4 and Table 4 show the state space and the corresponding probabilities of one of the models where death is modelled alongside disability. Figure 4 is similar to the MMM with recursive VAR(1) in Figure A2 (or Figure 2). Panel a of Table 4, the transition matrix for morbidity, is almost the same as Panel a in Table A2 (or Table 2). Transitions between morbidity states rely on morbidity and disability at time t. The major difference is in Panel b of Table 4 (the transition matrix for disability), where disability at time $t+1$ is only dependent on morbidity and disability at time t (for example, Pr ($disability_{t+1} = \text{DF} | morbidity_t = \text{MF}, disability_t = \text{DF}, \mathbf{X}$) is $\gamma_{mf.1}$), instead of being dependent on disability at time t and morbidity at time t and $t+1$.

(Figure 4 about here)

The probabilities in Table 4 also have a close relationship with the transition probabilities in Figure 3, but we cannot reverse engineer all the probabilities in Figure 3 like we did in Table 3 because the relationship between contemporaneous changes in the two dimensions of health (or time-varying variables) is approximated. For example, an individual with $morbidity_t = \text{MF}$ and $disability_t = \text{DF}$ will have the same probability of transitioning

tianyu.shen@anu.edu.au

to $disability_{t+1} = \text{DF } or \text{ D}$ regardless of their morbidity state at time $t + 1$. Conceptually speaking, this is not a bad approximation. People are most likely to be diagnosed with morbidities at the early stage of the disease with mild symptoms. The impact of a given disease accumulates over the life course and may lead to higher chances of being disabled at older ages. Most of the effects of chronic disease are accumulative rather than the result of an instant shock to the body (Chou et al. 2021), hence morbidity and disability onset are unlikely to be contemporaneous. However, this assumption may not hold true with other variables. For example, if the other event is not morbidity but instead a measure of whether the individual had experienced a fall, there would likely be a very strong concurrent relationship between falls and disability. Additionally, the accumulative effect would be very small. In other words, a person who experienced a fall and was not disabled in the first year may indicate that they were not severely injured and are not more likely to be disabled in the future years due to the fall. Therefore, like all other modelling, it is important to justify the selection of model with the empirical and theoretical evidence.

(Table 4 about here)

A comparison of results from the complex multistate method and two types of MMM (*recursive* and *reduced-form*) is presented in Table 5 and Figures A3, A4 & A5. Health expectancies are calculated through three models: complex five-state, MMM with recursive VAR(1) and MMM with *reduced-form* VAR(1) are compared in Table 5. Inside the parentheses are 95% confidence intervals from bootstrapping. For brevity, we only show the results at age 70 from cohort 1934-1943 (the full comparison for all age groups can be found in Figures A3 & A4 in the Appendix). The three models have very close point estimates for all

tianyu.shen@anu.edu.au

expectancies, with differences under 0.1 for all expectancies. Based on the confidence interval, none of them is significantly different from one another. In Figures A3 & A4, the MMM models are compared to complex multistate model. The figures are superimposed on each other so that it is easier to examine the difference. Figure A5 further disaggregates results in Figure A4 by initial morbidity status, which is a type of status-base life expectancy. Similar to Table 5, none of the age and gender groups are significantly different.

In short, the benefit of the alternative model is apparent. The dependent variables have fewer categories because they are now estimating the joint probability. Fewer categories mean larger sample sizes in each category and more reliable estimates, especially if a given transition is rare in some groups. The model is simpler, with fewer interaction terms. Nevertheless, this alternative model has additional assumptions. Whether these assumptions are supported can be understood and gauged through empirical evidence, allowing the researcher to decide whether it is an acceptable trade-off.

Example 2: marital status and healthy life expectancy

As a second example to demonstrate the flexibility of MMM, we apply the MMM to model self-rated health and marital status. Each outcome has three categories, leading to a 10-state state space which is quite large to estimate as a complex multistate model. Furthermore, marital status and health are very different domains of the life course. Even though there are papers (Goldman et al. 1995; Jia and Lubetkin 2020; Rendall et al 2011) about the association between marital status and health/survival, the mechanism is likely to be indirect through behavior change (Wilson and Oswald 2005), social support (Becker et al. 2019; Berkman 1984), and other long-term accumulated processes (Verbrugge 1979). Thus, it is a good case to demonstrate the advantage of MMM when the contemporaneous relationship between outcomes is theorized to be weak.

tianyu.shen@anu.edu.au

(Table 5 about here)

Marriage selection theory suggests that healthier individuals are more likely to get married (Goldman 1993; Murray 2000). In our analyses below, we exclude the never-married group. We also hypothesize that, conditional on being ever married, the current health of the respondent is likely not a strong predictor of marital status but not the other way around. We deploy this hypothesis partly because it is conceptually plausible and partly because we want to demonstrate the flexibility of the MMM. Therefore, this hypothesis can be removed depending on the research question. The regression models are constructed as follows,

$$\text{logit}(marital_{t+1}) = \alpha_m + \beta_{m1} marital_t + \beta_{m2} age_t + \beta_{m3} Sex, \quad (8)$$

$$\text{logit}(health_{t+1}) = \alpha_h + \beta_{h1} marital_t + \beta_{h2} health_t + \beta_{h3} age_t + \beta_{h4} Sex. \quad (9)$$

The transitions and state space of the model are illustrated in Figure 5, noting that the transition to death could be estimated in either equation. We also limit the state space to not allow direct transitions between divorced and widowed.

Since marital status varies over time, the population-averaged expectancies at any age are rather difficult to interpret and understand. Instead, we group individuals based on the period in the life-course that a marital status change occurs to explore the potential health impacts of a marital dissolution (i.e., divorced or widowed) on remaining healthy life expectancy, and how these may change over age. Table 6 presents the results of the remaining healthy life expectancy by gender and age according to the timing of a change in marital status. The first group of people remains married from age 55 to the starting age of remaining life expectancy and the other group experience at least once marital dissolution within a certain age

tianyu.shen@anu.edu.au

range (also including people who changing back to married within that age range). For example, a married man at age 55 who remained married at 64 could expect 7.39 years healthy life expectancy and total of 19.11 years of total remaining life expectancy remaining at age 65. In contrast, a married man at age 55 who experienced a marital dissolution between 55 and 64 could expect to live only 6.12 years of healthy life and 17.75 years of total remaining life at age 65.

(Figure 5 about here)

In general, individuals staying married over the period have higher remaining life expectancy. The two groups are also significantly different in their remaining healthy life expectancy at ages 65 and 75, and in the percent of remaining life lived in good health. However, this beneficial effect on health and survival diminishes with increasing age. Though insignificant, individuals who remain married at 85 have slightly higher healthy life expectancy compared to those who experience a marital dissolution between 75 and 84. As expected, women's healthy life expectancy and total life expectancy are always higher than men's at the same age.

(Table 6 about here)

Discussion and Conclusion

This paper introduces and develops a flexible method, the multiple multistate method (MMM), to estimate health expectancy in models with more than one time-varying variable. Previously,

tianyu.shen@anu.edu.au

time-varying variables other than the main health indicator would either be assumed static in the health expectancy estimation or be incorporated into the state space. Neither of these methods has been widely used, as they both come with substantial drawbacks: the static assumption may not be realistic in many cases, and the sample size required to estimate the complex multistate state space is larger than available in most longitudinal data sources. In addition, the multistate model is used in transition between labor force status (Hayward and Lichter 1998; Studer, Struffolino and, Fasang 2018), marital status (Schoen & Canudas-Romo 2006; Willekens et al. 1982; Zeng et al. 2012) and migration (Land and Rogers 1982; Raymer, Willekens, and Rogers 2019). Thus, the method is not confined to health expectancy, and it could be used to explore other durational expectancies based on the multistate model. Our approach opens new research directions using complex state spaces that are unfeasible to explore using the standard complex multistate model.

The MMM can fully reproduce the complex multistate model, but the advantage of MMM lies in its flexibility to trade off reductions in interaction terms for greater complexity in the modeled state space. As shown in the first example, the MMM with reduced interactions produces very similar results as compared to the complex multistate model. Furthermore, the second example also presents coherent findings with other related studies. Our results provide similar evidence on the protective effect of marriage on survival and health that is suggested in Rendall et al. (2011) and Jia and Lubetkin (2020). The protection effect also fades over age as found in the other studies. Robards et al. (2012) suggest that when it comes to the elderly other time-varying variables may also be important such as living arrangements, which is highly correlated with marital status.

With the MMM, it is feasible to generalize our framework and apply it to estimate more than two time-varying variables at a time. In general, it is recommended to keep the number of variables in a VAR small and only include variables that are plausibly related to each other

tianyu.shen@anu.edu.au

based on theory and empirical evidence (Stock and Watson 2020). This is particularly relevant for high-frequency time-series data where it is common to include multiple lags, because it would result in a large number of VAR parameters to be estimated (e.g., in Stock and Watson (2001), a VAR model with 9 variables and 4 lags would have 333 unknown coefficients to be estimated). However, this is less of an issue with annual data and in the context of demography where it is common to include only one lag. An alternative way to estimate a large number of parameters in VAR models with more than three variables is to impose a common structure on the coefficients (Stock and Watson 2001). For example, Bayesian methods have been introduced by Litterman (1986) to model six variables and by Sims and Zha (1998) to model as many as twenty variables in a VAR framework. The MMM approach could also potentially be combined with Bayesian Multistate Life Table Methods (Lynch & Zang 2022) to address very complicated research questions with many time-varying variables and relatively large state space in each of the multistate model.

As is common in statistical modeling, the reduced complexity of the MMM approach does come with a stronger set of assumptions than the complex multistate model. By providing a toolkit to flexibly reduce interaction terms, the MMM method substantially expands the complexity of multistate models that can be estimated using longitudinal sample survey data. Reducing these interaction terms can most clearly have an impact in cases where the two time-varying variables of interest have a strong contemporaneous relationship—that is, where a change in one variable has a strong, immediate impact on the likelihood of a change in the other variable(s). However, what interactions to include, and what to drop, is a question that must be largely guided by theory and previous evidence.

There are other limitations (or assumptions) related to multistate method. The multistate life table is essentially a discrete-time Markov process. One of the Markov properties is that it is a memoryless system, where the immediate next state only depends on the current state. This

tianyu.shen@anu.edu.au

is a common limitation of studies using MSLT with left-censored survey data. Cai, Schenker and Lubitz (2006) combined the semi-Markov model with backwards simulation algorithm to impute the starting point and avoid left-censored issue. This is a promising method to relax the Markov assumption by incorporating duration dependence but has not been widely used due to the small size and short follow-up period of most social surveys. Another limitation lies in the discrete-time approach and assuming no unobserved transitions between time points. The main reason for this approach is that HRS (and many other health surveys) are conducted about every two years. Cai et al. (2010) and Lynch and Zang (2022) also adopted this approach because of the similar limitation in the data. A recent study by Dudel and Schneider (2021) presented a way to quantify the potential bias from this assumption. Thus, it is important to bear these biases in mind when using the MMM and interpreting the results.

In conclusion, the MMM provides researchers with a powerful tool to estimate health expectancy with more than one time-varying variable (two and beyond) and in complex state spaces. Although these types of expectancy-based models are most common in estimating health expectancies, our approach could be used to explore a number of other durational expectancies such as time in employment, homelessness, and marriage. Overall, the MMM represents a flexible approach to estimating durational expectancies in complex models based on longitudinal sample survey data, and one that makes a wider array of social research questions possible in the multistate framework.

tianyu.shen@anu.edu.au

References

Authors. 2022. *Disability and morbidity among US birth cohorts, 1998-2018: a multidimensional test of dynamic equilibrium theory*. Working Paper (under review)

Agresti, A. 2007. *An introduction to categorical data analysis*. Hoboken, NJ: Wiley-Interscience.

Allison, P. D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13:61.

Becker, C., I. Kirchmaier, and S. T. Trautmann. 2019. "Marriage, parenthood and social network: Subjective well-being and mental health in old age." *PLOS ONE* 14:e0218704.

Berkman, L. F. 1984. "Assessing the Physical Health Effects of Social Networks and Social Support." *Annual Review of Public Health* 5:413-432.

Cai, L., M. Hayward, Y. Saito, J. Lubitz, A. Hagedorn, and E. Crimmins. 2010. "Estimation of multi-state life table functions and their variability from complex survey data using the SPACE Program." *Demographic Research* 22:129-158.

Cai, L. and J. Lubitz. 2007. "Was There Compression of Disability for Older Americans from 1992 to 2003?" *Demography* 44:479-495.

Cai, L., N. Schenker, and J. Lubitz. 2006. "Analysis of functional status transitions by using a semi-Markov process model in the presence of left-censored spells." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 55:477-491.

Canova, F. and M. Ciccarelli. 2013. "Panel Vector Autoregressive Models: A Survey." Pp. 205-246 in *VAR models in macroeconomics–new developments and applications: Essays in honor of Christopher A. Sims*, vol. 32: Emerald Group Publishing Limited.

tianyu.shen@anu.edu.au

Chang, L. and Y. Shi. 2021. "Mortality Forecasting with A Spatially Penalized Smoothed VAR Model." *ASTIN Bulletin* 51:161-189.

Chiu, C.-T. 2019. "Living arrangements and disability-free life expectancy in the United States." *PLOS ONE* 14:e0211894.

Chou, C.-Y., C.-J. Chiu, C.-M. Chang, C.-H. Wu, F.-H. Lu, J.-S. Wu, and Y.-C. Yang. 2021. "Disease-related disability burden: a comparison of seven chronic conditions in middle-aged and older adults." *BMC Geriatrics* 21.

Crimmins, E. M. 2004. "Trends in the Health of the Elderly." *Annual Review of Public Health* 25:79-98.

Dudel, C. 2021. "Expanding the Markov Chain Toolbox: Distributions of Occupation Times and Waiting Times." *Sociological Methods & Research* 50:401-428.

Dudel, C. and D. C. Schneider. 2021. "How bad could it be? Worst-case bounds on bias in multistate models due to unobserved transitions." *Sociological Methods & Research*:004912412199554.

Enders, W. 2004. *Applied econometric time series*. Hoboken, NJ: J. Wiley.

Epskam, S. 2013. "Estimation of (logistic) vector-autoregression models using generalized linear modelling." in *IMPS 2013*. Arnhem, the Netherlands.

Fullerton, A. S. and J. Xu. 2018. "Constrained and Unconstrained Partial Adjacent Category Logit Models for Ordinal Response Variables." *Sociological Methods & Research* 47:169-206.

Goldman, N. 1993. "Marriage Selection and Mortality Patterns: Inferences and Fallacies." *Demography* 30:189-208.

tianyu.shen@anu.edu.au

Goldman, N., S. Korenman, and R. Weinstein. 1995. "Marital status and health among the elderly." *Social Science & Medicine* 40:1717-1730.

Guibert, Q., O. Lopez, and P. Piette. 2019. "Forecasting mortality rate improvements with a high-dimensional VAR." *Insurance, mathematics & economics* 88:255-272.

Hayward, M. D. and D. T. Lichter. 1998. "A Life Cycle Model of Labor Force Inequality: Extending Clogg's Life Table Approach." *Sociological Methods & Research* 26:487-510.

Health and Retirement Study. 2021. "RAND HRS Longitudinal File 2018 (V1)." edited by U. o. Michigan. Ann Arbor, MI: University of Michigan.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56:1371-1395.

Huang, G., F. Guo, and G. Chen. 2021. "Multidimensional healthy life expectancy of the older population in China." *Population Studies* 75:421-442.

Huang, M. Y., R. R. Rojas, and P. D. Convery. 2020. "Forecasting stock market movements using Google Trend searches." *Empirical economics* 59:2821-2839.

Jia, H. and E. I. Lubetkin. 2020. "Life expectancy and active life expectancy by marital status among older U.S. adults: Results from the U.S. Medicare Health Outcome Survey (HOS)." *SSM - Population Health* 12:100642.

Kim, S. and J.-W. Lee. 2008. "Demographic changes, saving, and current account: An analysis based on a panel VAR model." *Japan and the world economy* 20:236-256.

Korinek, K., Z. Zimmer, and D. Gu. 2011. "Transitions in Marital Status and Functional Health and Patterns of Intergenerational Coresidence Among China's Elderly

tianyu.shen@anu.edu.au

Population." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66B:260-270.

Laditka, S. B. and D. A. Wolf. 1998. "New Methods for Analyzing Active Life Expectancy." *Journal of Aging and Health* 10:214-241.

Land, K. C. and A. Rogers. 1982. *Multidimensional mathematical demography*. New York: Academic Press.

Li, H. and Y. Lu. 2017. "Coherent forecasting of mortality rates: A sparse Vector-Autoregression approach." *ASTIN bulletin* 47:563-600.

Li, H. and Y. Shi. 2021. "Mortality Forecasting with an Age-Coherent Sparse VAR Model." *Risks (Basel)* 9:1-19.

Lièvre, A., N. Brouard, and C. Heathcote. 2003. "The Estimation of Health Expectancies from Cross-Longitudinal Surveys." *Mathematical Population Studies* 10:211-248.

Lim, W., G. Khemka, D. Pitt, and B. Browne. 2019. "A method for calculating the implied no-recovery three-state transition matrix using observable population mortality incidence and disability prevalence rates among the elderly." *Journal of Population Research* 36:245-282.

Litterman, R. B. 1986. "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience." *Journal of Business & Economic Statistics* 4:25-38.

Liu, Z., L. Han, Q. Feng, M. E. Dupre, D. Gu, H. G. Allore, T. M. Gill, and C. F. Payne. 2019. "Are China's oldest-old living longer with less disability? A longitudinal modeling analysis of birth cohorts born 10 years apart." *BMC medicine* 17:23-23.

tianyu.shen@anu.edu.au

Lynch, S. M. and J. S. Brown. 2005. "A New Approach to Estimating Life Tables with Covariates and Constructing Interval Estimates of Life Table Quantities." *Sociological Methodology* 35:189-237.

Lynch, S. M. and E. Zang. 2022. "Bayesian Multistate Life Table Methods for Large and Complex State Spaces: Development and Illustration of a New Method." *Sociological Methodology* 52:254-286.

Martikainen, P., H. Moustgaard, E. Einiö, and M. Murphy. 2014. "Life Expectancy in Long-Term Institutional Care by Marital Status: Multistate Life Table Estimates for Older Finnish Men and Women." *The Journals of Gerontology: Series B* 69:303-310.

Murray, J. E. 2000. "Marital protection and marital selection: Evidence from a historical-prospective sample of American men." *Demography* 37:511-521.

Payne, C. F. 2022. "Expansion, Compression, Neither, Both? Divergent Patterns in healthy, disability-free, and morbidity-free life expectancy across US birth cohorts, 1998-2016." *Demography* 59.

Raymer, J., F. Willekens, and A. Rogers. 2019. "Spatial demography: A unifying core and agenda for further research." *Population, Space and Place* 25:e2179.

Rendall, M. S., M. M. Weden, M. M. Favreault, and H. Waldron. 2011. "The Protective Effect of Marriage for Survival: A Review and Update." *Demography* 48:481-506.

Robards, J., M. Evandrou, J. Falkingham, and A. Vlachantoni. 2012. "Marital status, health and mortality." *Maturitas* 73:295-299.

Saito, Y., J.-M. Robine, and E. M. Crimmins. 2014. "The methods and materials of health expectancy." *Statistical journal of the IAOS* 30:209-223.

tianyu.shen@anu.edu.au

Schoen, R. and V. Canudas-Romo. 2006. "Timing Effects on Divorce: 20th Century Experience in the United States." *Journal of Marriage and Family* 68:749-758.

Sims, C. A. and T. Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* 39:949-968.

Sonnega, A., J. D. Faul, M. B. Ofstedal, K. M. Langa, J. W. Phillips, and D. R. Weir. 2014. "Cohort Profile: the Health and Retirement Study (HRS)." *International Journal of Epidemiology* 43:576-585.

Stock, J. H. and M. W. Watson. 2001. "Vector Autoregressions." *The Journal of Economic Perspectives* 15:101-115.

———. 2020. *Introduction to econometrics*. Harlow, England: Pearson.

Studer, M., E. Struffolino, and A. E. Fasang. 2018. "Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure." *Sociological Methodology* 48:103-135.

Verbrugge, L. M. 1979. "Marital Status and Health." *Journal of Marriage and Family* 41:267-285.

Willekens, F. J., I. Shah, J. M. Shah, and P. Ramachandran. 1982. "Multi-state Analysis of Marital Status Life Tables: Theory and Application." *Population Studies* 36:129-144.

Wooldridge, J. M. 2020. *Introductory econometrics: a modern approach*. Boston, MA: Cengage.

Yang, Z. and A. G. Hall. 2007. "The Financial Burden of Overweight and Obesity among Elderly Americans: The Dynamics of Weight, Longevity, and Health Care Cost." *Health Services Research* 43:849-868.

tianyu.shen@anu.edu.au

Yong, V. and Y. Saito. 2012. "Are There Education Differentials in Disability and Mortality Transitions and Active Life Expectancy Among Japanese Older Adults? Findings From a 10-Year Prospective Cohort Study." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 67B:343-353.

Zeng, Y., S. P. Morgan, Z. Wang, D. Gu, and C. Yang. 2012. "A Multistate Life Table Analysis of Union Regimes in the United States: Trends and Racial Differentials, 1970–2002." *Population Research and Policy Review* 31:207-234.

tianyu.shen@anu.edu.au

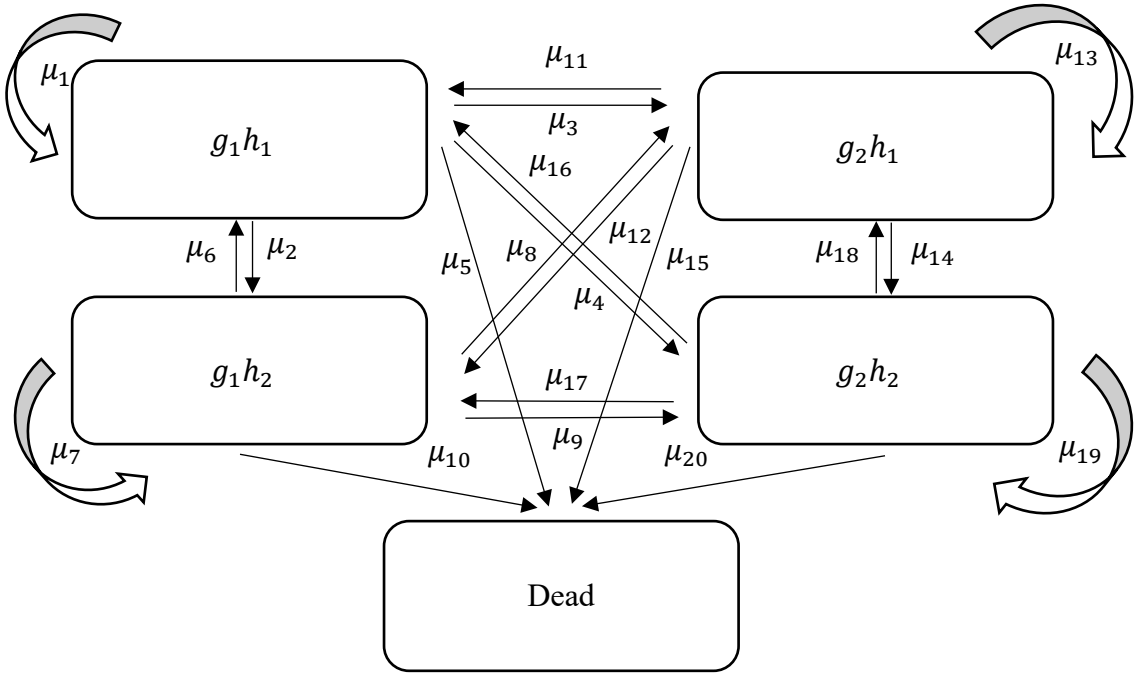

Figure 1. Complex multistate model

tianyu.shen@anu.edu.au

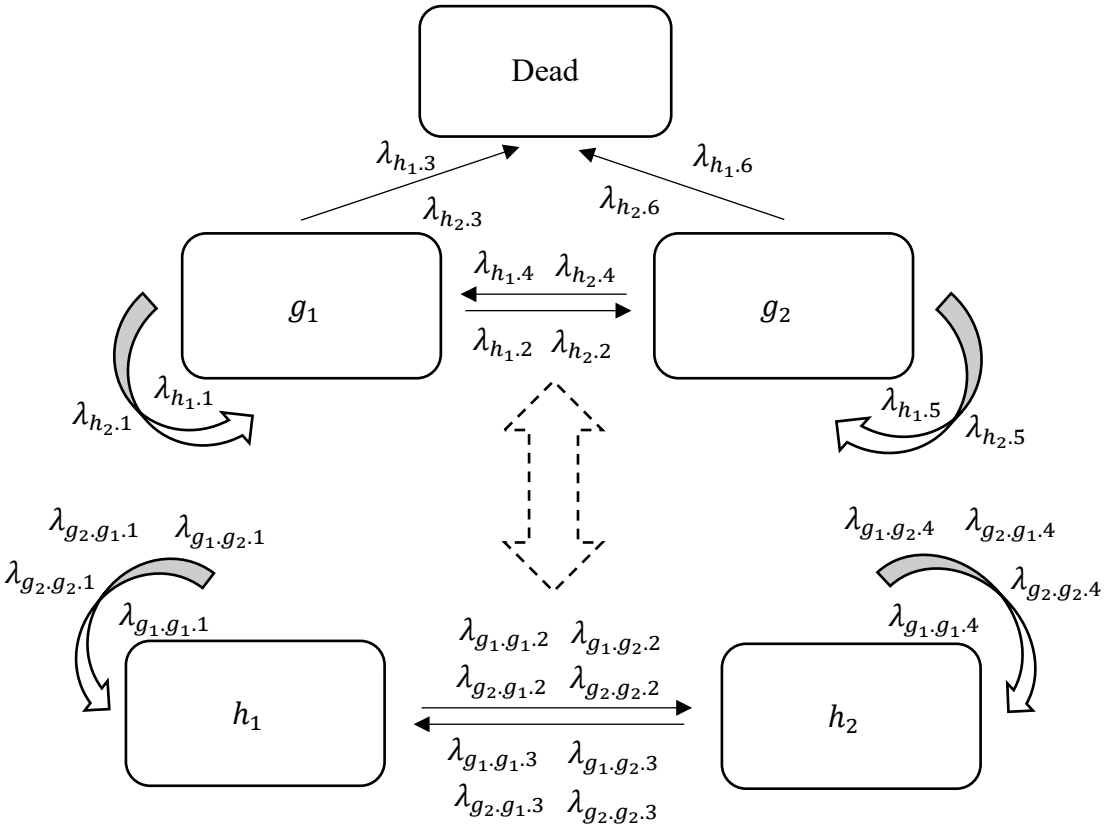

Figure 2. Multiple Multistate Method with bivariate *recursive* VAR(1) (Type a)

tianyu.shen@anu.edu.au

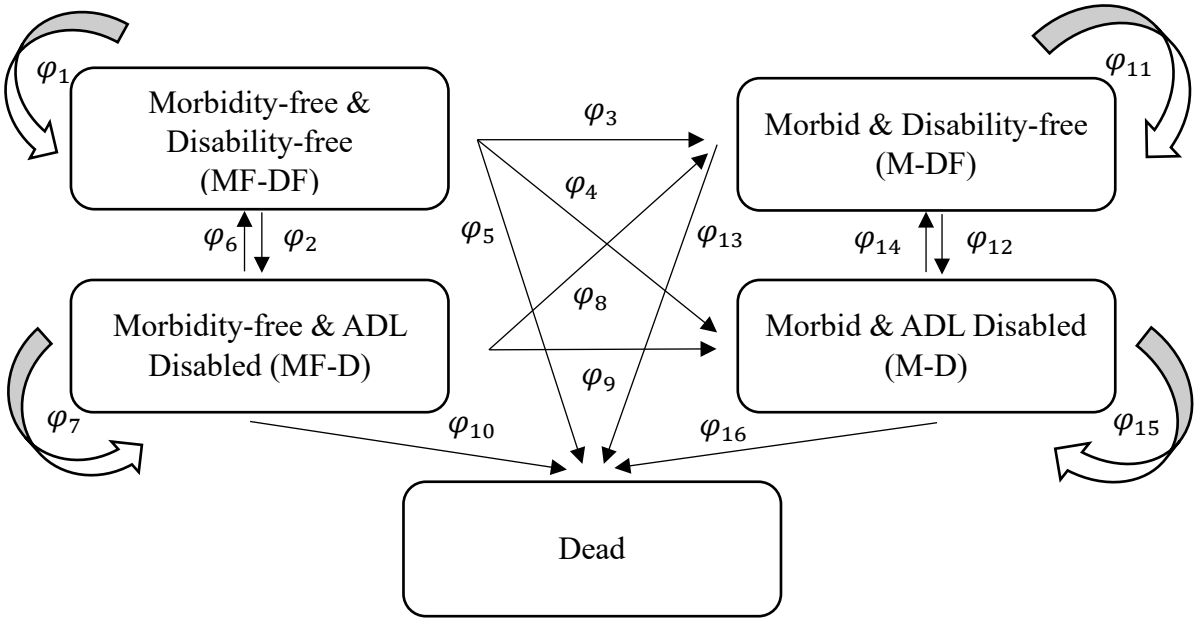

Figure 3. Complex Multistate Model with disability and morbidity

tianyu.shen@anu.edu.au

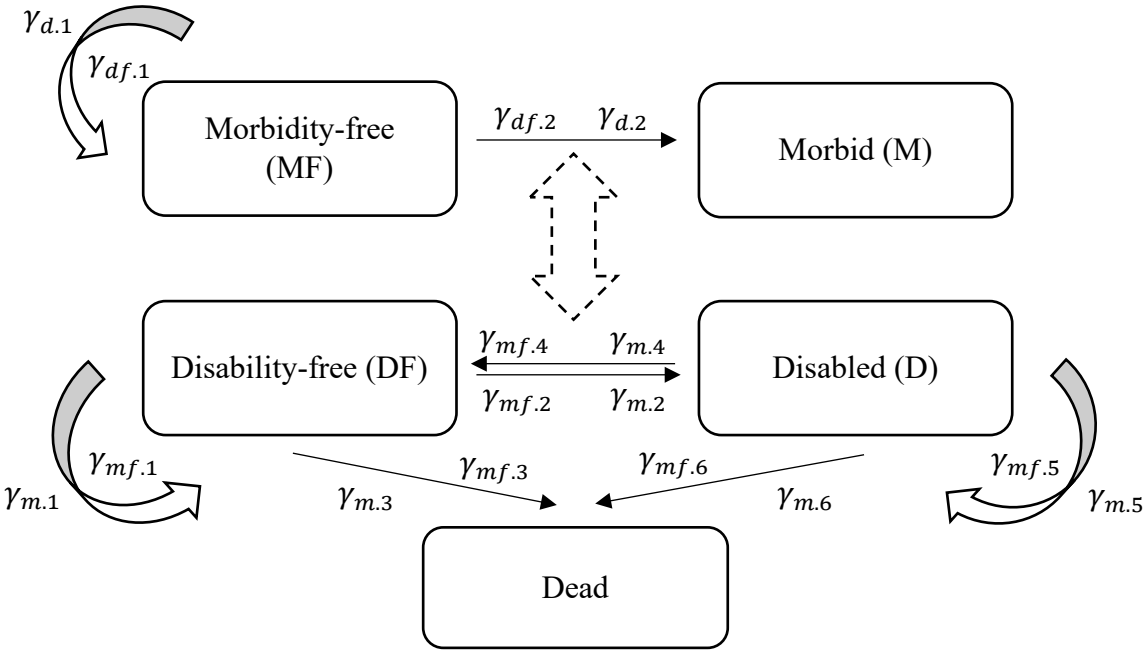

Figure 4. Multiple Multistate Method with *reduced-form* VAR(1)

tianyu.shen@anu.edu.au

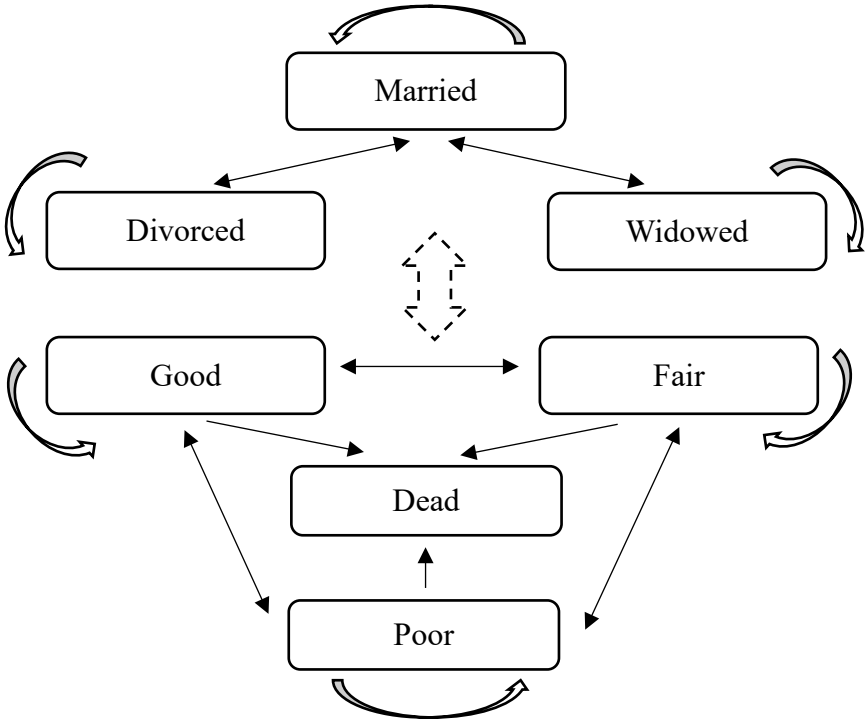

Figure 5. Marital status and health with MMM

tianyu.shen@anu.edu.au

Table 1. Transition probabilities of the Complex Multistate Model in Figure 1

	g_1h_1	g_1h_2	g_2h_1	g_2h_2	Dead
g_1h_1	μ_1	μ_2	μ_3	μ_4	μ_5
g_1h_2	μ_6	μ_7	μ_8	μ_9	μ_{10}
g_2h_1	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{15}
g_2h_2	μ_{16}	μ_{17}	μ_{18}	μ_{19}	μ_{20}

tianyu.shen@anu.edu.au

Table 2. Transition Probabilities of the Multiple Multistate Method in Figure 2

a. Transition Matrix for Variable *G*

	g_1	g_2	Dead
$g_{1_{h_1}}$	$\lambda_{h_1.1}$	$\lambda_{h_1.2}$	$\lambda_{h_1.3}$
$g_{2_{h_1}}$	$\lambda_{h_1.4}$	$\lambda_{h_1.5}$	$\lambda_{h_1.6}$
$g_{1_{h_2}}$	$\lambda_{h_2.1}$	$\lambda_{h_2.2}$	$\lambda_{h_2.3}$
$g_{2_{h_2}}$	$\lambda_{h_2.4}$	$\lambda_{h_2.5}$	$\lambda_{h_2.6}$

b. Transition Matrix for Variable *H*

	h_1	h_2
$h_{1_{g_1.g_1}}$	$\lambda_{g_1.g_1.1}$	$\lambda_{g_1.g_1.2}$
$h_{2_{g_1.g_1}}$	$\lambda_{g_1.g_1.3}$	$\lambda_{g_1.g_1.4}$
$h_{1_{g_1.g_2}}$	$\lambda_{g_1.g_2.1}$	$\lambda_{g_1.g_2.2}$
$h_{2_{g_1.g_2}}$	$\lambda_{g_1.g_2.3}$	$\lambda_{g_1.g_2.4}$
$h_{1_{g_2.g_1}}$	$\lambda_{g_2.g_1.1}$	$\lambda_{g_2.g_1.2}$
$h_{2_{g_2.g_1}}$	$\lambda_{g_2.g_1.3}$	$\lambda_{g_2.g_1.4}$
$h_{1_{g_2.g_2}}$	$\lambda_{g_2.g_2.1}$	$\lambda_{g_2.g_2.2}$
$h_{2_{g_2.g_2}}$	$\lambda_{g_2.g_2.3}$	$\lambda_{g_2.g_2.4}$

tianyu.shen@anu.edu.au

Table 3. Equivalent probabilities between the complex multistate model and MMM

Table 2 Panel a	Equivalent in Table 1	Table 2 Panel b	Equivalent in Table 1
$\lambda_{h_1.1}$	$\mu_1 + \mu_2$	$\lambda_{g_1.g_1.1}$	$\frac{\mu_1}{\mu_1 + \mu_2}$
$\lambda_{h_1.2}$	$\mu_3 + \mu_4$	$\lambda_{g_1.g_1.2}$	$\frac{\mu_2}{\mu_1 + \mu_2}$
$\lambda_{h_1.3}$	μ_5	$\lambda_{g_1.g_1.3}$	$\frac{\mu_6}{\mu_6 + \mu_7}$
$\lambda_{h_1.4}$	$\mu_{11} + \mu_{12}$	$\lambda_{g_1.g_1.4}$	$\frac{\mu_7}{\mu_6 + \mu_7}$
$\lambda_{h_1.5}$	$\mu_{13} + \mu_{14}$	$\lambda_{g_1.g_2.1}$	$\frac{\mu_3}{\mu_3 + \mu_4}$
$\lambda_{h_1.6}$	μ_{15}	$\lambda_{g_1.g_2.2}$	$\frac{\mu_4}{\mu_3 + \mu_4}$
$\lambda_{h_2.1}$	$\mu_6 + \mu_7$	$\lambda_{g_1.g_2.3}$	$\frac{\mu_8}{\mu_8 + \mu_9}$
$\lambda_{h_2.2}$	$\mu_8 + \mu_9$	$\lambda_{g_1.g_2.4}$	$\frac{\mu_9}{\mu_8 + \mu_9}$
$\lambda_{h_2.3}$	μ_{10}	$\lambda_{g_2.g_1.1}$	$\frac{\mu_{11}}{\mu_{11} + \mu_{12}}$
$\lambda_{h_2.4}$	$\mu_{16} + \mu_{17}$	$\lambda_{g_2.g_1.2}$	$\frac{\mu_{12}}{\mu_{11} + \mu_{12}}$
$\lambda_{h_2.5}$	$\mu_{18} + \mu_{19}$	$\lambda_{g_2.g_1.3}$	$\frac{\mu_{16}}{\mu_{16} + \mu_{17}}$
$\lambda_{h_2.6}$	μ_{20}	$\lambda_{g_2.g_1.4}$	$\frac{\mu_{16}}{\mu_{16} + \mu_{17}}$
		$\lambda_{g_2.g_2.1}$	$\frac{\mu_{13}}{\mu_{13} + \mu_{14}}$
		$\lambda_{g_2.g_2.2}$	$\frac{\mu_{14}}{\mu_{13} + \mu_{14}}$
		$\lambda_{g_2.g_2.3}$	$\frac{\mu_{18}}{\mu_{18} + \mu_{19}}$
		$\lambda_{g_2.g_2.4}$	$\frac{\mu_{19}}{\mu_{18} + \mu_{19}}$

tianyu.shen@anu.edu.au

Table 4. Transition Probabilities of the Multiple Multistate Method in Figure 4

a. Transition Matrix for Morbidity

	MF	M
MF_{df}	$\gamma_{df.1}$	$\gamma_{df.2}$
M_{df}	0	1
MF_{d}	$\gamma_{d.1}$	$\gamma_{d.2}$
M_{d}	0	1

b. Transition Matrix for Disability

	DF	D	Dead
DF_{mf}	$\gamma_{mf.1}$	$\gamma_{mf.2}$	$\gamma_{mf.3}$
D_{mf}	$\gamma_{mf.4}$	$\gamma_{mf.5}$	$\gamma_{mf.6}$
DF_{m}	$\gamma_{m.1}$	$\gamma_{m.2}$	$\gamma_{m.3}$
D_{m}	$\gamma_{m.4}$	$\gamma_{m.5}$	$\gamma_{m.6}$

tianyu.shen@anu.edu.au

Table 5. Comparison of three models

Gender	Health State	Traditional five-state	MMM (*recursive*)	MMM (*reduced*)
Men	MF-DF	2.54 (2.37, 2.73)	2.52 (2.36, 2.68)	2.50 (2.32, 2.65)
	MF-D	0.24 (0.20, 0.30)	0.25 (0.20, 0.30)	0.27 (0.23, 0.31)
	M-DF	4.82 (4.63, 5.00)	4.83 (4.64, 5.00)	4.87 (4.70, 5.05)
	M-D	0.96 (0.90, 1.07)	1.00 (0.91, 1.08)	0.95 (0.87, 1.03)
	Dead	1.34 (1.25, 1.41)	1.33 (1.26, 1.43)	1.33 (1.25, 1.42)
Women	MF-DF	3.29 (3.11, 3.45)	3.24 (3.10, 3.43)	3.23 (3.05, 3.38)
	MF-D	0.33 (0.29, 0.39)	0.33 (0.29, 0.38)	0.37 (0.34, 0.42)
	M-DF	4.02 (3.86, 4.18)	4.05 (3.89, 4.20)	4.08 (3.95, 4.25)
	M-D	1.28 (1.19, 1.39)	1.31 (1.22, 1.40)	1.25 (1.16, 1.35)
	Dead	0.99 (0.91, 1.06)	0.99 (0.92, 1.05)	0.98 (0.91, 1.05)

tianyu.shen@anu.edu.au

Table 6. Remaining healthy life expectancy by marital status

Gender	Starting age (Age range)	Health state	Remain married	Changed to non-married
Men	65 (55-64)	Good	**7.39 (7.09, 7.70)**	**6.12 (5.73, 6.52)**
		Fair	6.63 (6.37, 6.88)	6.19 (5.88, 6.53)
		Poor	5.09 (4.84, 5.33)	5.43 (5.12, 5.78)
		Total	**19.11 (18.70, 19.51)**	**17.75 (17.20, 18.32)**
	75 (65-74)	Good	**4.31 (4.05, 4.56)**	**3.52 (3.26, 3.82)**
		Fair	4.20 (3.98, 4.39)	3.83 (3.59, 4.07)
		Poor	3.61 (3.40, 3.82)	3.81 (3.53, 4.08)
		Total	**12.11 (11.78, 12.45)**	**11.17 (10.78, 11.58)**
	85 (75-84)	Good	2.22 (2.01, 2.45)	1.89 (1.68, 2.11)
		Fair	2.29 (2.09, 2.47)	2.14 (1.96, 2.35)
		Poor	2.24 (2.05, 2.45)	2.41 (2.18, 2.65)
		Total	6.76 (6.46, 7.05)	6.45 (6.14, 6.75)
Women	65 (55-64)	Good	**8.89 (8.51, 9.26)**	**7.50 (7.09, 7.86)**
		Fair	7.24 (6.99, 7.52)	6.88 (6.58, 7.16)
		Poor	5.87 (5.64, 6.13)	6.37 (6.08, 6.70)
		Total	**22.02 (21.61, 22.39)**	**20.75 (20.27, 21.15)**
	75 (65-74)	Good	**5.28 (5.00, 5.57)**	**4.45 (4.19, 4.70)**
		Fair	4.87 (4.64, 5.08)	4.55 (4.33, 4.74)
		Poor	4.29 (4.09, 4.52)	4.59 (4.37, 4.84)
		Total	**14.44 (14.11, 14.76)**	**13.59 (13.24, 13.92)**
	85 (75-84)	Good	2.74 (2.51, 2.99)	2.38 (2.18, 2.57)
		Fair	2.83 (2.64, 3.02)	2.70 (2.53, 2.86)
		Poor	2.77 (2.56, 2.98)	2.99 (2.81, 3.20)
		Total	8.34 (8.04, 8.62)	8.06 (7.82, 8.35)

tianyu.shen@anu.edu.au