

【保险精算】

# 基于非齐次 Markov 模型的 长期护理保险定价研究

张琳 汤薇

**【摘要】**本文运用四状态时间非齐次马尔科夫模型对 65~105 岁老年人的健康转移强度建模,在 Tweedie-GLM 模型的框架下,加入年龄效应、时间趋势和年龄—时间交互效应。实证结果表明,除了常见的年龄效应外,时间趋势和年龄—时间交互作用也是解释健康状态转移的重要因素。本文还基于精算理论预测预期寿命和护理需求期望时间,并给出 2020 年的长期护理保险精算费率表,结果发现在各个时刻下的女性,相较于相同年龄和初始健康状态下的男性都具有绝对的生存优势;虽然在 2002~2012 年男性具有健康优势,但随着时间的推移,男性在失能状态的持续时间将超过女性,而女性在受损状态的持续时间一直稳步超过男性,二者叠加使得 2020 年初始状态为失能的男性长期护理保险费率显著高于女性。

**【关键词】**非齐次 Markov;健康转移概率;失能;长期护理保险

**【作者简介】**张琳,湖南大学风险管理与保险精算研究所所长、金融与统计学院教授, E-mail: lindazhang0203@126.com;汤薇,上海财经大学金融学院博士。

**【原文出处】**《保险研究》(京),2020.7.108~121

## 一、引言与文献综述

人口老龄化是包括中国在内的全球许多国家面临的重大挑战。根据联合国标准,65 周岁人口数量占总人口比例为 7% 时,该国家或地区进入老龄化社会。我国大约在 2000 年左右进入老龄化社会。截至 2018 年底,我国 65 周岁及以上人口 1.66 亿,占总人口的 11.9%。<sup>①</sup>虽然我国不是老龄化程度最严重的国家,但老龄化速度却是世界上最快的国家之一。长期护理是进入老龄化的国家必须面临的难题。城乡老年人生活状况抽样调查结果显示,2015 年全国城乡失能、半失能老年人口的总量约为 4063 万,相比于 2013 年的 3750 万人,增长速度约 4.1%。中国将在接下来的几十年中保持对健康服务的高度需求,而家庭规模结构的核心化、小型化以及女性地位的提高等,导致传统的代际支持模式正逐步退化,家庭已难以支撑老年人的长期照护需求。因而政府也在积极促进商业性养老金融产品的发展来分担健康服务成本的负担,包括商业医疗

保险,商业长期护理保险(Long Term Care Insurance, LTCI)和住房反向抵押养老保险产品等(Li et al., 2017)。研究老年人健康状态的演变,可以为政府制定相关政策提供建议。对老年人失能时间的研究是建立长期照护保险制度、失能老人补贴制度和制定家庭养老支持政策的基础(张立龙和张翼, 2017),合理准确的精算费率厘定是商业 LTCI 定价的基础。

本文着眼于时间趋势给老年人健康状态演变以及长期护理保险定价带来的影响,为推广商业长期护理保险产品提供参考和依据。近年来关于长期护理保险方面的研究成果已日益丰富。国外的研究主要集中在融资、需求和可保性三个领域(Eling and Ghavibazoo, 2019),国内关于个人长期护理保险的研究主要涉及长期护理保险的需求预测、影响因素分析以及定价研究等。其中,定价研究主要是以 Markov 过程的实现和健康状态转移矩阵的构建为基础,而转移概率矩阵的计算不仅是长期护理

保险定价的精算基础,也是预测未来护理状态人口规模和健康生存时间的基础。国内学者的研究一般基于时齐性 Markov 过程。时齐性假设是指转移强度为不变常数,转移概率只与时间间隔有关,而与所处的时间点无关。如黄枫和吴纯杰(2012)、胡宏伟等(2015)、王新军和王佳宇(2018)等均利用 logit 模型刻画老年人的健康状态转移概率,将年龄、城乡和婚姻状态等作为自变量考虑进来。或者利用年龄非齐次 Markov 过程,如胡晓宁等(2016)基于广义线性模型(GLM)估计随年龄变化的健康转移强度,并利用三状态 Kolmogorov 微分方程计算转移概率,得到不同状态下的老年人口预期生存时间和不同初始健康状态下的精算现值。荆涛等(2016)运用通过市场调研法获得的转移概率矩阵,对北京市政策性长期护理保险进行定价。崔晓东(2017)基于中国老年人健康长寿影响因素调查(CLHLS)数据库本身构建原始转移概率矩阵,提出转移强度为分段常数、按年龄队列矩阵相乘的 Markov 方法。

只有少数论文考虑时间非齐次的 Markov 过程。在时间非齐性 Markov 框架下,转移概率与所处的时间点有关,但与在时间点持续的时间无关。而中国的人口变化发生得非常快,因此在健康状态转移中考虑时间效应是很重要的。Hanewald et al. (2017)首次将年龄效应、时间趋势和年龄—时间交互效应加入 GLM 框架下对不考虑恢复率的三状态 Markov 过程建模。Li et al. (2017)提出了一个多状态潜在因子强度模型来对转移率和恢复率进行建模,其中时间项和随机项分别用来捕捉确定趋势和不确定性。而在 GLM 的框架下,Lally and Hartman (2016)通过对大型商业长期护理保险公司的死亡率、失效率和索赔率数据的实证研究发现,相对于零膨胀泊松、负二项和 Tweedie 误差的广义线性和可加模型,Tweedie - GLM 能够提供更准确的预测。张连增和谢厚谊(2017)认为只要 Tweedie - GLM 法的估计结果与观测数据相吻合,Tweedie - GLM 法因待估参数较少,从而总是优于 SPG (Separate Poisson - Gamma) 法。

近年来国外关于 Semi - Markov 过程的文章开

始涌现。Semi - Markov 过程是指转移概率不仅与所处的时间点有关,且与在时间点持续的时间也有关。Biessy (2015)基于法国长期护理公共援助数据给出了四状态 Semi - Markov 模型的构建过程,并利用蒙特卡洛方法将该模型应用于新的长期护理保险产品定价。Biessy (2016)又利用 Semi - Markov 模型研究了失能的潜在病理对 LTC 寿命的影响。Fuino and Wagner (2018)基于 Semi - Markov 模型发现转移概率在性别、年龄、护理水平和护理类型中所花费的时间存在显著差异。刘乐平等(2020)首次运用 Semi - Markov 模型,结合国内的健康与养老追踪调查(CHARLS)和瑞典的状态停留时间数据,测算出我国中老年人口的状态转移概率。

综上,由于国内缺乏跟踪记录失能状态改变的数据库,且国外数据库是否适用还值得商榷,国内的研究主要集中在齐次 Markov 假设下的 Logit 模型或年龄非齐次 Markov 假设下的 GLM 模型等,尚未有学者跳出传统 GLM 模型对时间非齐次的 Markov 健康转移过程进行运算研究。另外,失能时间的预测和精算费率的计算大多基于一次跟踪调查数据,对政策的制定和保险产品的定价缺乏实用性和前瞻性。最后,根据年龄时间非齐次 Markov 得到的结果是多维的,因而在矩阵运算上要求较高。因此,结合我国失能现状的特点,本文在四状态年龄和时间非齐次 Markov 框架下,运用加入了时间趋势效应的 Tweedie - GLM 模型对 CLHLS (2002 ~ 2014 年)四次健康转移跟踪调查数据建模,对不同年龄、时刻和初始健康状态下的男性和女性老人预期寿命以及状态持续时间进行横截面和时间序列两个维度的分析,并给出 2020 年长期护理保险价格表。

## 二、数据及其描述统计

### (一)数据

本文使用中国老年健康影响因素跟踪调查(CLHLS)纵向调查数据库的样本。该调查分别于 1998 年、2000 年、2002 年、2005 年、2008 年、2011 年、2014 年在全国 22 个省(自治区、直辖市)展开,这些地区的人口约占全国总人口的 85%。本文以 2002 ~ 2005 年、2005 ~ 2008 年、2008 ~ 2011 年、2011 ~ 2014 年四次调查样本中 65 ~ 105 岁的老年人为研

究对象。<sup>②</sup>CLHLS 从基本情况、性格特征、认知能力、生活方式、日常活动能力、个人背景及家庭结构和生理健康六个维度,以 180 多项问题访问了受访者当时的健康状况。CLHLS 是国际上对“年龄最大的老人(80 岁以上)”的最大纵向调查数据库,且经过众多学者的论证,发现依据 CLHLS 数据库计算出来的死亡率和发病率数据具有良好的质量(曾毅,2013)。

## (二) 界定长期护理状态

在老年人长期护理保障制度的设计中,转移概率的估计对“长期护理失能”的定义很敏感,且如何界定护理状态将直接影响老年人的获益资格和受益水平。这影响了商业和政策性长期护理保险计划对给付触发器的设计(Fong et al., 2015)。本文对护理状态的界定采用学术界广泛使用的 ADL 独立性评判准则,即反映老年人活动能力的 6 个日常项目(洗澡、穿衣、室内活动、如厕、吃饭和大小便)。另外,调查也采集了器具性日常活动能力(IADLs)的八项指标,包括到邻居家串门、做饭、独自外出买东西、洗衣、乘坐交通工具等。由于这些指标反映的都是老年人在社区中独立生活所需的较高级的技能,考虑到中国老年人普遍的经济水平和文化程度,本文将受访者无法完成全部八项 IADLs 才记为健康受损。

在此基础上,认知能力严重受损也被广泛纳为界定长期护理状态的一项重要指标,而 CLHLS 数据库关于认知能力的调查包括两个方面的问题:一是根据认知功能简易量表(MMSE)设置的关于反应能力、注意力和理解能力、回忆以及语言和自我协调能力的问题;二是关于是否患有以下精神和与记忆相关的疾病(帕金森、痴呆、癫痫等)。因为关于 MMSE 的测试问题容易出现虚假作答的情况,因此本文采用王新军和王佳宇(2018)关于认知功能障碍的定义,若老年人样本中存在着医院鉴定的情感、精神以及与记忆相关的疾病,则认定为存在认知功能障碍。综上,本文将老人健康状态划分为四类,具体见表 1。

## (三) 描述统计

本文通过删除健康状态等关键变量缺失或后

续无法跟进的观测值,对数据进行清理,以避免潜在的异常值。接着对样本的转移进行描述性分析,发现样本的高频健康转移主要集中在 65 ~ 105 岁的区间上,所以剔除小于 65 岁以及大于 105 岁的老人样本。Fuino and Wagner (2018) 通过将每个性别一年龄组合分解为同质模型,来分别计算各状态之间的转换概率,结果发现男性和女性之间的转换概率存在显著差异。黄匡时和陆杰华(2014)发现在平均预期照护时间上城乡差异不明显。

表 1 老年人健康状态分类情况

状态名称	代码	状态描述
健康	H	无任何功能障碍
受损	M	8 项 IADLs 障碍或 1 ~ 2 项 ADLs 障碍; 无认知功能障碍
失能	S	3 项及以上 ADLs 障碍或认知功能障碍
死亡	D	身故

因此,本文仅按性别(男性和女性)来拆分数据集以解决非同质性问题。在模型估计中,将 2002 年定义为  $t = 1$ ,并相应地将每次调查时间设置为  $t = (1, 4, 7, 10)$ ,用来反映影响转移强度的因素是每次调查之间的时间间隔。表 2 是四次调查的原始转移次数,总共观察到 25615 次健康转移,总的恢复率为 6.5%。<sup>③</sup>

为了计算健康、受损和失能状态下样本人群的风险暴露程度,本文使用调查中的精确的调查时间和死亡日期,当确切日期缺失时,使用调查月份的第 15 天。并假设健康状态之间的转移发生在每次调查的中点。表 3 给出了原始风险暴露年,风险暴露总年数为 104712,男性和女性分别占样本的 44.4% 和 55.6%。

通过给定年龄和时间的健康转移次数除以相应状态的风险暴露年得到图 1 和图 2(见下页)表示的原始转移强度热图。图 2 中的空白区域表示原始数据中不存在相应状态的风险暴露。图 1 和图 2 的纵轴均代表年龄,横轴表示调查的时间区间,右边的细柱状条表示原始转移强度的大小刻度。图 1 和图 2 显示,健康到死亡 H - D 和失能到死亡 S - D 的最大刻度分别为 0.5 和 1.5,明显高于 H - M 和 M - S 的 0.16,说明相较于其他状态之间的转移,死

表 2 健康转移次数

时间	H→M		H→S		H→D		M→H		M→S		M→D		S→H		S→M		S→D	
	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女
2002~2005年	285	42	169	273	1333	1424	146	276	79	188	595	1181	43	70	14	49	332	838
2005~2008年	208	340	217	310	1363	1453	100	220	61	121	396	800	28	54	14	49	293	741
2008~2011年	423	605	227	351	1488	1670	69	113	70	192	384	870	15	43	20	33	282	707
2011~2014年	216	313	118	152	655	623	90	122	54	131	259	402	21	30	17	38	248	459
占比	9.5%		7.1%		39.1%		4.4%		3.5%		19.1%		1.2%		0.9%		15.2%	

表 3 风险暴露年

时间	State H		State M		State S		总计
	男	女	男	女	男	女	
2002~2005年	11146	11638	1618	3388	669	1624	30083
2005~2008年	11229	12038	1094	2397	553	1491	28803
2008~2011年	11452	12406	962	2288	498	1326	28932
2011~2014年	7025	7062	812	1447	496	1004	17846

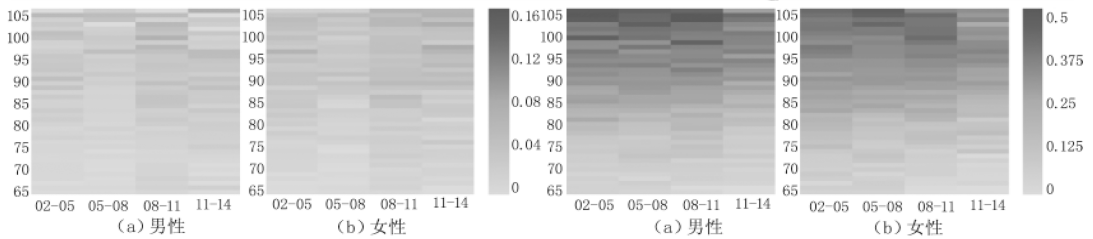


图 1 从健康到受损 H-M、健康到死亡 H-D 的原始转移强度

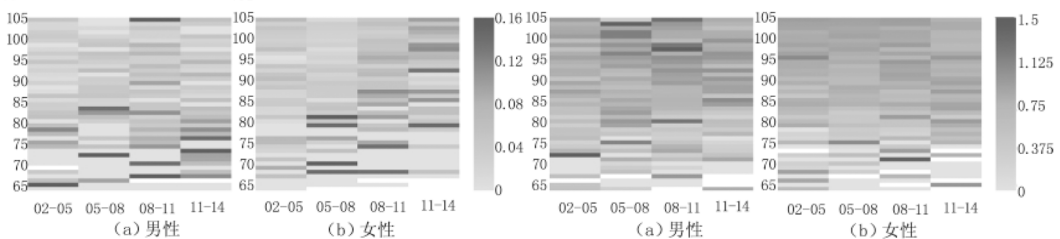


图 2 从受损到失能 M-S、失能到死亡 S-D 的原始转移强度

亡转移过程发生得更为频繁。在图 1 中,可以看出随着调查时间的推移,颜色程度有轻微的加深,但在 M-S 或 S-D 中,可能不存在时间趋势,且不同年龄的个体之间转移强度差异较大。

### 三、基于非齐次 Markov 过程的转移概率模型

#### (一) 四状态时间非齐次 Markov 模型

假定每个人的健康转移过程服从四状态 Markov 过程,即未来状态的条件概率过程仅仅取决于当

前所处的状态,与过去的状态独立。图 3 展示了四状态 Markov 过程包括的三个转移状态(健康、受损和失能)和一个吸收状态(死亡)。

本文假设所有的 9 个健康转移都遵循时间非齐性的 Markov 过程,其中转移概率依赖于转移发生的时刻:

$$\begin{aligned}
 P_{ij}(x, t, h) &= \Pr(S(x+h, t+h) = j | S(x, t) = i) \\
 u_{ij}(x, t) &= \lim_{h \rightarrow 0^+} P_{ij}(x, t, h) / h \quad (1)
 \end{aligned}$$



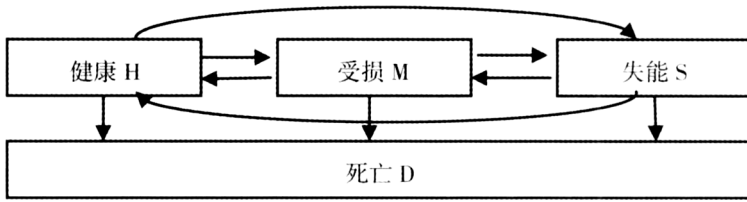


图3 四状态 Markov 过程

其中,  $x$  代表年龄,  $t$  代表时间,  $h \geq 0$ ,  $S(x, t)$  表示个体在年龄  $x$  和时间  $t$  的随机健康状态,  $i, j \in \{H, M, S, D\}$ 。  $P_{ij}(x, t, h)$  表示在年龄  $x$  和时间  $t$  处于状态  $i$  到年龄  $x+h$  和时间  $t+h$  到状态  $j$  的转移概率,  $u_{ij}(x, t)$  表示年龄  $x$  和时间  $t$  的转移强度。

(二) 健康转移概率的估计模型

本文利用 Tweedie 复合泊松分布模型估计随年龄和时间变化的健康状态转移强度  $u_{ij}(x, t)$ , 并根据反映模型拟合程度的 AIC 准则选择最优模型形式。 Tweedie 类分布是指数分布族中一类特殊的概率分布, 其中  $E(u_{ij}(x, t)) = \lambda$ ,  $Var(u_{ij}(x, t)) = \phi V(\mu) = \phi \mu^p$ ,  $\phi > 0$  为离散参数,  $p$  为尺度参数, 在复合泊松分布中,  $1 < p < 2$ 。 将均值  $\mu = E(u_{ij}(x, t))$  定义为通过对数连接函数  $\eta$  的线性预测函数形式, 即  $\eta(\mu) = \log(\mu) = X\beta = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 xt + \beta_4 x^2 + \beta_5 x^2 t$ 。

基于健康状态转移强度  $u_{ij}(x, t)$  的估计值, 本文运用四状态 Kolmogorov 微分方程计算相应的健康状态转移概率  $P_{ij}(x, t, h)$ , 公式如下:

$$\frac{d}{dt} P_{ij}(x, t, h) = \sum_{k \neq 0}^n [P_{ik}(x, t, h) u_{kj}(x, t, h) - P_{ij}(x, t, h) u_{jk}(x, t, h)] \quad (2)$$

从状态  $i$  进入状态  $j$  的瞬时概率等于从任何其他状态进入状态  $j$  的概率, 减去离开状态  $j$  到另一个状态的概率。 使用 Euler 方法求解方程, 即在  $h = 0$  的初始时刻有:

$$P_{ik}(x, t, 0) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

选择一个足够小的持续期  $h$ , 上述微分方程近似得到:

$$P_{ij}(x, t, h) \approx P_{ij}(x, t, 0) + h \sum_{k \neq 0}^n [P_{ik}(x, t, h) u_{kj}(x, t, h) - P_{ij}(x, t, h) u_{jk}(x, t, h)] \quad (3)$$

对于时间非齐次 Markov 模型, 每一个年龄  $x$ 、初始时刻  $t$  和持续期  $h$  都有一个转移强度矩阵。

(三) 各健康状态下老年人的预期生存时间

由于时间非齐次 Markov 模型假定不同年龄的人群在未来相等的时间内健康状态变化是不相同的。 为方便计算, 本文假设持续期  $h = 1$  年, 即转移强度或转移概率在 1 年内为常数, 同时状态的变化在 1 年内只考虑期初和期末时的状态, 中间的互相

转化状态不计, 令  $Q(x, t, 0) = \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix}$  表示在时刻  $t$

年龄为  $x$  岁的个体初始健康状态的分布, 令  $M(x, t, h)$  表示年龄为  $x$  岁的个体从时刻  $t+h$  到  $t+h+1$  的健康状态转移概率矩阵 ( $4 \times 4$ ), 则年龄为  $x$  岁的人在时刻  $t$  到  $t+h$  ( $h \geq 1$ ) 时的健康状态分布可表示为:

$$\begin{bmatrix} Q_H(x, t, h) \\ Q_M(x, t, h) \\ Q_S(x, t, h) \end{bmatrix} = \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix} \prod_{j=1}^h M(x, t, h) = \begin{bmatrix} q_{HH}(x, t, h) q_{HM}(x, t, h) q_{HS}(x, t, h) q_{HD}(x, t, h) \\ q_{MH}(x, t, h) q_{MM}(x, t, h) q_{MS}(x, t, h) q_{MD}(x, t, h) \\ q_{SH}(x, t, h) q_{SM}(x, t, h) q_{SS}(x, t, h) q_{SD}(x, t, h) \end{bmatrix} \quad (4)$$

式中  $q_{ij}(x, t, h)$  表示年龄为  $x$  岁的个体在时刻  $t$  时初始健康状态为  $i \in \{H, M, S\}$  在  $t+h$  时刻的健康状态为  $j \in \{H, M, S, D\}$  的概率, 此时状态持续的期望时间为  $E[T(x, t)]$ , 具体公式:

$$E_{ij}(T(x, t)) = \sum_{h=0}^{w-x} q_{ij}(x, t, h) x \in [65, 105] \quad (5)$$

其中,  $w$  为极限年龄, 本文设定为 105 岁, 即所有个体达到 105 岁时均死亡。 式(5)适用于计算任何年龄和任何初始状态下的预期生存时间和在不同状态下的持续时间。

(四) 长期护理保险精算费率

本文通过计算不同年龄和不同时刻的老龄人口的健康状态转移概率,结合定价公式,得到长期护理保险的价格。由上文的定义可知, $P_{ij}(x, t, h)$ 表示在年龄  $x$  和时间  $t$  时处于状态  $i \in \{H, M, S\}$  经过  $h$  时刻到状态  $j \in \{H, M, S, D\}$  的转移概率。假设折现因子为  $v$ ,  $b_i$  表示在状态  $i$  时保险公司给付的金额,因此在  $t$  时刻年龄为  $x$  岁的被保险人在状态  $i$  时期望获得的长期护理保险赔付  $B_i(x, t, h)$  表示为:

$$B_i(x, t, h) = b_i + \sum_j v \times P_{ij}(x, t, h + 1) \times B_j(x, t, h + 1) \quad (6)$$

根据式(6)可以求出任意  $t$  时刻  $x$  岁的被保险人期望获得的赔付金额。

四、转移概率矩阵的实证分析

(一) 健康状态转移概率矩阵

本文运用 R 软件中的 `cplm` 包对 Tweedie 复合泊松分布模型进行估计,运用 AIC 准则选取模型最优形式,分别得到了男性和女性的 9 个健康状态转移强度,进一步根据拟合得到的数据计算校正拟合优度(adjusted  $R^2$ )。并与一般 GLM 模型拟合得到的 AIC 结果进行对比,以此来评价 Tweedie 复合泊

松分布模型的拟合效果。表 4 和表 5 分别给出了最优模型的参数估计结果。

从 AIC 的结果来看, Tweedie - GLM 整体上要优于一般 GLM 的拟合效果。不同转移强度之间 adjusted  $R^2$  的差异较大,可能与某些状态下发生转移的样本量过少有关。从表 4 和表 5 的  $\beta_1$  系数可以看出 M - H、S - H、S - M 均呈现出随年龄增加恢复率变小的负年龄效应;男性个体 S - D 除年龄和年龄的平方项不显著外,其余系数均显著,意味着男性老人从失能到死亡的转移强度因时间和年龄的不同存在较大差异,且从系数大小来看,相比于女性个体,男性个体的失能预期寿命可能会随着时间的推移出现较大波动(见图 6)。 $\beta_2$  系数显示 H - M、M - S(男性)、H - D、M - D 和 S - D 具有明显的时间趋势效应,系数为正表示随着时间的推移,转移强度增加,而不同的期初状态转移到死亡的强度均表现出负时间效应,表明随着时间的推移,死亡率减少,预期寿命变长;女性老年人从健康和受损状态转移到死亡的系数大于男性,表明女性的生存时间可能高于男性,但男性从失能转移到死亡的强度随着时间的推移在减少,意味着未来男性在失能状态的持续时间可能会长于女性。

表 4 男性转移强度的参数估计

系数	H - M	H - S	H - D	M - H	M - S	M - D	S - H	S - M	S - D
$\beta_0$	-21.72***	-21.14***	-18.46***	2.99***	-2.98***	-4.21***		17.66*	-0.596***
$\beta_1 (\times 10)$	3.781***	3.722***	3.065***	-0.614***		0.3506***	-0.3467***	-4.756*	
$\beta_2 (\times 10^2)$	3.473**		-2.269***		4.679 <sup>-</sup>	-109.8*			-212.0**
$\beta_3 (\times 10^3)$						25.43*			4.346*
$\beta_4 (\times 10^3)$	-1.928***	-1.970***	-1.311***					2.644*	
$\beta_5 (\times 10^4)$						-1.457*			-2.190*
AIC	-808.98	-842.18	-685.50	-103.38	-159.09	-73.78	54.02	83.58	125.06
AIC(GLM)	-834.76	-932.28	-541.95	-195.06	-365.85	-192.18	-253.29	-279.78	156.60
adjusted $R^2$	0.436	0.232	0.900	0.366	0.012	0.470	0.249	0.044	0.173
离散参数	0.0184	0.0150	0.0119	0.1848	0.1165	0.1147	0.2890	0.4637	0.2090
尺度参数	1.2042	1.1562	1.2414	1.2989	1.1903	1.0443	1.2743	1.3402	1.0423

注:线性预测函数为  $\beta_1 + \beta_1 x + \beta_2 t + \beta_3 xt + \beta_4 x^2 + \beta_5 x^2 t$ 。\*\*\*表示  $P < 0.001$ , \*\*表示  $P < 0.01$ , \*表示  $P < 0.05$ , <sup>-</sup>表示  $P < 0.1$ , AIC 表示模型的拟合程度,绝对值越小,表明模型拟合得越好。

表 5 女性转移强度的参数估计

系数	H - M	H - S	H - D	M - H	M - S	M - D	S - H	S - M	S - D
$\beta_0$	-20.58***	-17.23***	-20.83***	2.79***	-16.52**	-5.03***	3.76***		-4.13***
$\beta_1 (\times 10)$	3.494***	2.690***	3.539***	-0.574***	3.032*	0.4176***	-0.7696***	-0.7747***	0.3649***
$\beta_2 (\times 10^2)$			-3.058***			-153.8***			-156.0*
$\beta_3 (\times 10^3)$						33.78***			35.81*
$\beta_4 (\times 10^3)$	-1.722***	-1.278***	-1.564***		-1.636*			0.4322**	
$\beta_5 (\times 10^4)$						-1.852***			-2.036*
AIC	-905.55	-915.43	-740.05	-210.71	-201.27	-175.85	-15.87	-54.88	47.27
AIC(GLM)	-891.63	-934.62	-620.59	-196.84	-447.68	-345.99	-159.63	-418.03	-82.19
adjusted R <sup>2</sup>	0.708	0.500	0.913	0.403	0.079	0.710	0.221	0.198	0.438
离散参数	0.0071	0.0096	0.0072	0.1931	0.0966	0.0566	0.4551	0.2684	0.1447
尺度参数	1.1029	1.1110	1.0810	1.4035	1.1737	1.0322	1.4633	1.3652	1.0442

注:同表 4。

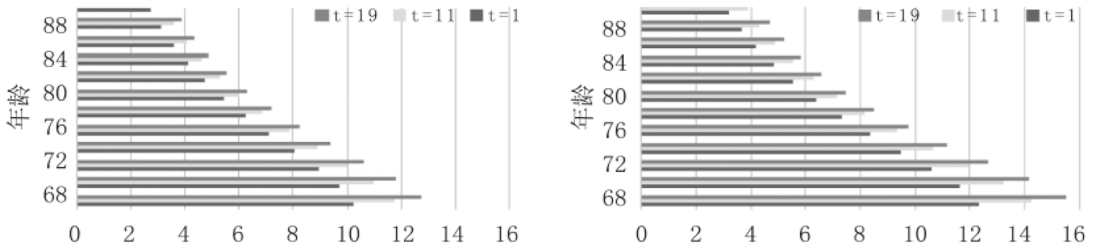


图 4 初始状态为健康的男性和女性个体在  $t=1, 11$  和  $19$  时刻的预期寿命(年)

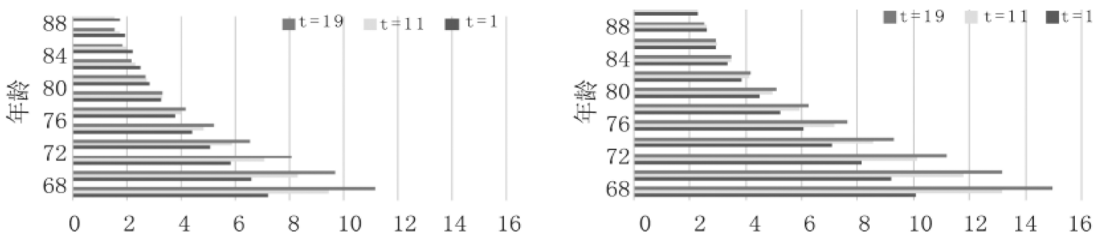


图 5 初始状态为受损的男性和女性个体在  $t=1, 11$  和  $19$  时刻的预期寿命(年)

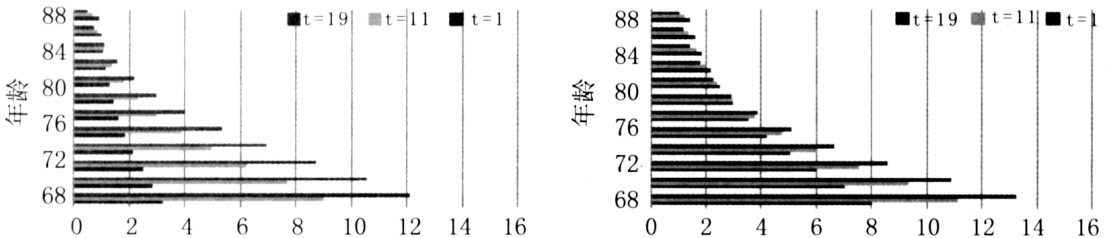


图 6 初始状态为失能的男性和女性个体在  $t=1, 11$  和  $19$  时刻的预期寿命(年)

$\beta_3$  系数表示年龄—时间的交互效应, M - D 和 S - D 的系数为正表示健康转移随着时间的推移总

体上有所恶化,且时间的恶化效果因年龄而异。从系数大小来看,女性的恶化程度要高于男性。从健

康 H 转移到其他状态 M、S 和 D 的  $\beta_4$  系数均为负,表示转移强度随着年龄的增长而增长,但增长的幅度下降。S - M (女性) 中虽然年龄的二次项为正,但年龄一次项的负效应会有一定程度的抵消,可能导致处于中间年龄段的个体恢复率较高。 $\beta_5$  系数仅在 M - D 和 S - D 中显著,表明对于年龄较大的老人来说其从受损或失能状态转移到死亡状态的速度更快,但时间趋势起到了一定的保护作用。

总体而言,年龄和时间因素是解释中国高龄老人健康转移模式的重要变量。此外,一些模型依赖于年龄—时间的相互作用。这种相互作用捕捉了随年龄不同而不同的时间趋势。根据估计的模型最优形式和参数值,代入式(3)可计算每个年龄在不同时刻的转移概率矩阵,然后进行预期寿命的估计和长期护理保险费率的计算。周海珍(2015)表示多状态 Markov 模型中涉及到大量的迭代运算和矩阵相乘,而保险公司目前使用的 Excel 和 Prophet 软件在矩阵方面的计算能力较差,因而在实务中尚未普及,但多状态 Markov 模型是长期护理保险定价的发展趋势。本文选取的是年龄和时间均为非齐次的多状态 Markov 模型,所以矩阵的运算量更为庞大。限于篇幅,本文仅计算  $t = 1$  (2002 年)、 $t = 11$  (2012 年) 和  $t = 19$  (2020 年) 三个时刻的预期生存时间和 2020 年的长期护理保险精算费率,以显示年龄、时间以及年龄和时间的交互效应的影响。

## (二) 预期寿命和状态持续时间

本文运用四状态 Kolmogorov 微分方程计算各年龄人群在  $t = 1$ 、11 和 19 三个时刻的健康状态转移概率矩阵,并进一步运用式(5)计算出各年龄人群在不同时刻的预期寿命和状态持续时间。限于篇幅,本文仅展现 68 ~ 90 岁的男性和女性老年人在不同时刻和不同初始健康状态的预期寿命以及不同时刻在受损和失能两个状态的持续时间。

### 1. 不同时刻和不同初始健康状态的预期寿命

#### (1) 横截面维度

整体上看,68 ~ 90 岁老年个体的预期寿命随年

龄的增加而逐渐减小,且女性的减小幅度快于男性,说明男性具有一定的健康优势。通过对不同初始健康状态个体的预期寿命进行分析,发现女性老人的预期寿命显著高于男性,生存优势明显。以  $t = 11$  时刻(2012 年)68 岁老人为例,期初为健康状态的女性预期寿命为 14.27 年,男性为 11.73 年,女性比男性长 2.54 年。这与张立龙和张翼(2017)研究的 2008 年 65 ~ 67 岁的女性老人余命比男性老人长 2.47 年的结论基本相符。

从健康和受损以及受损和失能之间预期寿命的差值来看,不同的时刻有不同的表现。如  $t = 1$  时刻(2002 年),年龄较小时,男性预期寿命的差值均显著大于女性。以 65 岁为例,男性老年人的预期寿命差分别为 3.02 年和 4.01 年,而女性则为 2.25 年和 2.09 年。但随着年龄的增加,不同健康初始状态对寿命的影响减弱,个体主要受到死亡因素的影响,因而到 90 岁时,男性老年人的预期寿命差缩减为 0.98 年和 0.84 年,女性老年人缩减为 0.91 年和 0.89 年。 $t = 11$  时刻(2012 年)时,男性和女性老年人的寿命的差值均表现出随年龄的增加而出现先增大后减小的非线性变化特征。男性老年人健康和受损预期寿命差值的极大值点在 76 岁,极大值约为 3 年;受损和失能预期寿命差值的极大值点在 78 岁,差值约为 1 年。而女性老年人寿命差的极大值点分别为 78 岁和 72 岁。这种先增加后减小的非线性变化差异可能与“选择性死亡”假说有关,即期初处于相对生存劣势的群体,经过死亡的选择效应,只有相对强健的个体才能生存下来,随着年龄的增加,差距逐渐缩小,意味着死亡风险减弱了不同健康初始状态对余命的影响。这种选择机制的影响表现在不同初始状态下的男女两个群体中的身体强健者和身体孱弱者之间。

#### (2) 时间维度

在  $t = 1$ 、11 和 19 时刻,女性在 68 ~ 90 岁之间的期望生存时间均高于男性,表明女性老人的生存优势不随时间的推移和初始健康状态的不同而改变,其期望寿命始终高于男性。另外,男性和女性老年人的预期寿命都表现出随时间的推移而增加的趋势,这与医疗水平的进步和生活条件的改善有



关。期初为受损状态的女性在预期寿命上表现出更大的优势;但期初为失能状态的男性在预期寿命上具有更强的波动性,其在  $t(1,11)$  之间增长较快,在  $t(11,19)$  之间增速放缓,这与时间效应在男性健康的动态转移过程中发挥了更大的作用有关。随着年龄的增加,男性和女性预期寿命的时间效应均被死亡因素削弱。另外,时间趋势对年纪较小的个体影响较大,期望寿命随年龄和时间趋势的变化而发生显著变化。

### 2. 不同时刻在不同健康状态的持续时间

老年人的护理需求与其在受损、失能状态持续的时间直接相关,从图7和图8可以看出,老年人的护理需求与年龄大小和时间趋势都呈现非线性关系。同一时刻不同年龄表现出金字塔形,在  $t=1$  时刻(2002年)男性和女性老年人在受损和失能状态的持续时间在不同的年龄上并没有显著差异,而随着时间的推移,各持续时间均随年龄的增加出现更大的坡度,表现为金字塔呈现扩宽趋势,特别是68~80岁的年龄段塔身扩宽程度十分明显。从图7可以看出,受损金字塔左侧呈现出逐渐放宽的趋势,在  $t=1,11$  和19时刻,不同年龄的女性老年人在受损状态持续时间均表现出比男性更快的增长,且显著高于男性,这与黄枫和吴纯杰(2012)利用CLHLS(2005~2008年)数据得到的女性老人在需

要护理状态停留的时间可能是男性老人2倍以上的结论基本相符。

但随着时间的推移,在  $t=11$  时刻(2012年),男性老人在80岁之前的护理时间就已经超过女性,在  $t=19$  时刻(2020年)时,失能金字塔的上部进一步收窄,下部进一步放宽,导致64~84岁之间的男性老人的护理时间显著高于女性。相较于其他学者的研究,本文发现虽然男性老人在  $t=1$  时刻(2002年)的失能状态持续时间较短,但随着时间的推移,男性在失能状态的持续时间将显著超过女性,说明男性的失能状态持续时间受时间趋势和年龄效应的影响较大,这将导致男性在未来的长期护理保险定价中具有较高的保险费率。

### (三) 老年人长期护理保险费率

胡宏伟等(2015)利用2015年“中国老年家庭与养老服务全面调查”数据进行加权得到全国范围的长期护理服务项目定价基准,其中,日常生活照料服务费用约6960元/年,医疗保健护理服务约3600元/年,精神慰藉约4500元/年。荆涛等(2016)通过对北京市政策性长期护理保险进行定价发现,一、二、三级护理分别约22266元/年、28932元/年和34314元/年。王新军和王佳宇(2018)假设折现率为3.5%,处于轻度、中度和重度时每年获得的保险赔付分别为10000元、20000元和30000元。

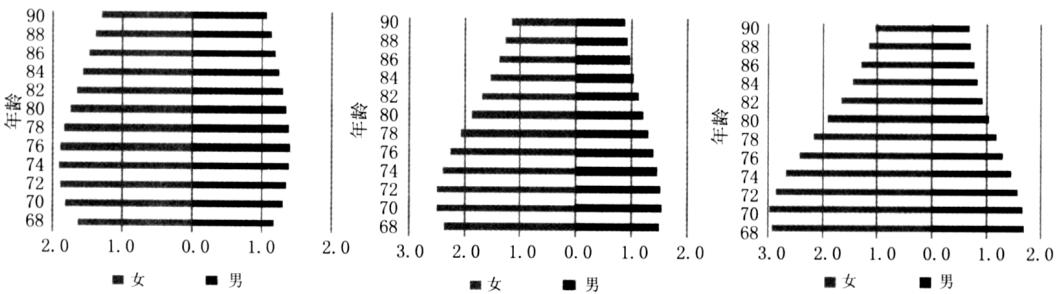


图7 男性和女性个体在  $t=1,11$  和19时刻在受损状态的持续时间(年)

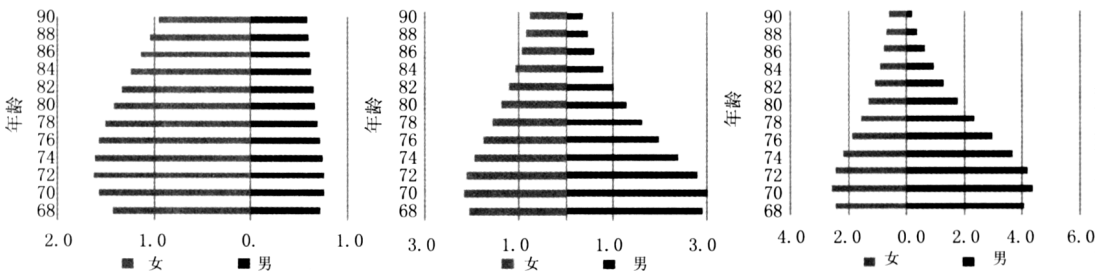


图8 男性和女性个体在  $t=1,11$  和19时刻在失能状态的持续时间(年)

综合考虑本文对失能状态的定义以及定价标准的基础适用性,本文最终假设被保险人处于受损状态时每年获得 10000 元的保险给付,处于失能状态时每年获得 20000 元的保险给付,且市场无风险利率为 3.5%,则每年的折现因子均为  $1/(1 + 3.5\%)$ 。通过不同年龄老年人在各个时刻估计的健康状态转移概率矩阵结果,代入公式(6)分别计算在  $t = 19$  时刻(2020 年)时男性和女性在不同初始年龄和健康状态下保障到 85 岁的长期护理保险费率,结果见表 6。

以初始状态为健康的被保险人为例,投保年龄为 70 岁的男性和女性的趸交纯保费分别为 17044.48 元和 18634.45 元,等价于老年人拥有 15 年的被保障年限,总的纯保险赔付成本分别为 17044.48 元和 18634.45 元。从整体结果来看,不同初始健康状态的男性和女性都表现出随着年龄的增长出现先增

长后下降的趋势,意味着 70 岁左右的老年人具有更高的长期护理保险需求。随着年龄的增加,趸交纯保费开始逐渐减少,具体的数值大小是保障期限缩短和长期护理风险增加共同作用的结果。初始健康状态的恶化导致男性和女性的长期护理保险费率均明显升高,且男性在不同初始健康状态的费率相比于女性波动幅度更大;当被保险人初始状态为健康时,女性的长期护理保险费率略高于男性,反映了男性和女性对长期护理保险的需求相差不多。而在受损状态时,不同年龄的男性和女性表现出不同的变动特征,在年龄相对较小时,男性的费率高于女性,而随着年龄的增长,女性的费率显著高于男性,意味着年龄较大的女性转移到失能状态的概率高于男性;当期初状态为失能时,男性在各个年龄下的护理费率水平均显著高于女性,且年龄效应的波动更明显,意味着男性在失能状态的停

表 6 65 ~ 80 岁的男性和女性老年人在  $t = 19$  时刻的长期护理保险费率表

年龄	健康 H		受损 M		失能 S	
	男性	女性	男性	女性	男性	女性
65	9942.70	10270.57	32499.74	29146.92	63692.46	49146.37
66	13482.04	13644.53	41032.81	36118.26	80895.50	59553.67
67	15361.65	15487.73	45889.36	40405.78	91105.75	65457.67
68	16478.89	16979.02	48793.75	43715.79	97439.71	69216.98
69	16931.99	17987.49	50171.86	46043.59	100765.99	71108.64
70	17044.48	18634.45	50464.17	47534.60	101706.86	71477.25
71	17033.55	19327.71	49950.37	48575.34	100651.15	70851.13
72	16855.95	19769.06	48708.75	48934.73	97808.76	69272.46
73	16510.20	20064.11	46886.70	48782.81	93530.84	67077.83
74	15992.95	20221.39	44659.09	48216.37	88253.37	64491.79
75	15309.53	20122.93	42210.86	47233.12	82433.62	61623.42
76	14526.74	19833.16	39721.83	45958.55	76460.19	58629.58
77	13719.51	19409.96	37310.81	44491.32	70573.01	55624.06
78	12933.19	18914.39	35036.23	42917.34	64900.65	52694.87
79	12184.77	18396.32	32924.72	41304.13	59523.00	49904.59
80	11486.23	17889.84	30991.29	39700.96	54495.61	47293.97

留时间较长,可以看出年龄、时间和年龄—时间的交互项在其中发挥较大作用。这与图 8 显示的失能状态持续时间是一致的,且通过表 4 的系数可以看出男性相较于女性随时间趋势、年龄—时间的交互项的变动幅度更大。

## 五、结论与展望

本文在综合考虑长期护理现状以及数据集特征的基础上,选取了 CLHLS 四次跟踪调查数据,并根据 ADL、IADLs 和认知能力将老年人的健康状态划分为健康、受损和失能。在年龄时间非齐次 Markov 框架下,利用 Tweedie - GLM 模型分别刻画 65 ~ 105 岁男性和女性老年人的健康转移强度,并基于 AIC 准则选取模型最优形式,以此为基础计算老年人未来健康的动态演变过程以及长期护理保险精算现值。本文的主要结论如下:

第一,整体来说,在男性和女性老年人的健康转移过程中,时间趋势的作用不可忽视。本文的模型也可以拓展到任意时刻、任意年龄的计算,但预测时间过长会损失一定的精度。另外,在未来样本数据更加丰富的条件下,可以根据其他可能的影响因素(如城乡、婚姻状况等)进一步细分子集,降低异质性。政府应建设失能状态转移跟踪数据库,以提高对失能持续时间预测的准确性。

第二,从横截面维度看,预期寿命随年龄的增加逐渐减小,且男性的减小幅度小于女性,意味着男性具有一定的健康优势;女性老人的预期寿命始终显著高于男性,生存优势明显。从时间维度看,老年人在不同初始健康状态之间的寿命差表现出随年龄的增加而先增大后减小的趋势。最后,虽然女性在受损状态的持续时间一直略高于男性,但男性在失能状态的持续时间随时间的推移增幅明显,未来有进一步超过女性的可能。

第三,初始状态为受损时,随着年龄的增加,老年人的长期护理保险费率表现出先增加后减小的趋势;初始状态为失能时,男性长期护理保险费率显著高于女性。随着时间的推移,未来的长期护理保险费率可能会进一步提高,但收入水平较高的家庭仍可以负担相应的费用,以此来缓解家庭照护者在工作与照料老人之间的冲突;而经济收入水平相

对较低的群体,可能需要更进一步的支持,如定点护理服务机构或家护、巡护服务等。

## 注释:

①数据来源:中国统计局。

②CLHLS 从 2002 调查开始,增加了年龄区间在 65 ~ 79 岁的样本。考虑到老人群体的代表性,本文仅研究 65 ~ 105 岁的老年人。

③所以考虑恢复率是很有必要的。

## 参考文献:

- [1] 崔晓东. 中国老年人口长期护理需求预测——基于多状态分段常数 Markov 分析[J]. 中国人口科学, 2017, (6): 82 - 93.
- [2] 胡宏伟, 李延宇, 张澜. 中国老年长期护理服务需求评估与预测[J]. 中国人口科学, 2015, (3): 79 - 89.
- [3] 胡晓宁, 陈秉正, 祝伟. 基于家庭微观数据的长期护理保险定价[J]. 保险研究, 2016, (4): 57 - 67.
- [4] 黄枫, 吴纯杰. 基于转移概率模型的老年人长期护理需求预测分析[J]. 经济研究, 2012, (s2): 119 - 130.
- [5] 黄匡时, 陆杰华. 中国老年人平均预期照料时间研究——基于生命表的考察[J]. 中国人口科学, 2014, (4): 92 - 101.
- [6] 荆涛, 杨舒, 谢桃方. 政策性长期护理保险定价研究——以北京市为例[J]. 保险研究, 2016, (9): 74 - 88.
- [7] 刘乐平, 唐爽, 程瑞华. 考虑状态停留时长的我国中老年人人口状态转移概率测算[J]. 保险研究, 2020, (2): 102 - 113.
- [8] 王新军, 王佳宇. 基于 Markov 模型的长期护理保险定价[J]. 保险研究, 2018, 366(10): 89 - 101.
- [9] 曾毅. 中国老年健康影响因素跟踪调查(1998 - 2012)及相关政策研究综述(下)[J]. 老龄科学研究, 2013, 1(2): 63 - 71.
- [10] 张立龙, 张翼. 中国老年人失能时间研究[J]. 中国人口科学, 2017, (6): 94 - 104.
- [11] 张连增, 谢厚谊. Tweedie 分布在车险费率厘定中的应用[J]. 保险研究, 2017, (1): 80 - 90.
- [12] 周海珍. 基于多状态 Markov 模型的长期护理保险定价研究. 2015 中国保险与风险管理国际年会论文集[C]. 2015: 9.
- [13] Biessy G. Long - Term Care Insurance: A Multi -

State Semi - Markov Model to Describe the Dependency Process for Elderly People[J]. Bulletin Francais d'Actuariat, 2015, 15(29) : 41 - 74.

[14] Biessy G. A Semi - Markov Model with Pathologies for Long - Term Care Insurance, Working Paper, Université d'Evry Val d'Essonne, 2016.

[15] Eling M, Ghavibazoo O. Research on Long - term Care Insurance: Status Quo and Directions for Future Research [J]. The Geneva Papers on Risk and Insurance - Issues and Practice, 2019, 44(2) : 303 - 356.

[16] Fong Joelle H, Adam W. Shao, and Michael Sherris. Multistate Actuarial Models of Functional Disability [J]. North American Actuarial Journal, 2015, 19(1) : 41 - 59.

[17] Fuino M. and Wagner J. Long - term Care Models and Dependence Probability Tables by Acuity Level: New Empirical Evidence from Switzerland [J]. Insurance: Mathematics and Economics, 2018, 81 : 51 - 70.

[18] Hanewald Katja, Han Li, and Adam Wenqiang Shao. Modeling Multi - State Health Transitions in China: A Generalized Linear Model with Time Trends [J]. Social Science Electronic Publishing, 2017.

[19] Lally Nathan R. and Brian M. Hartman. Predictive Modeling in Long - Term Care Insurance [J]. North American Actuarial Journal, 2016, 20(2) : 1 - 24.

[20] Li Han, Katja Hanewald, and Shang Wu. Estimating Healthy Life Expectancy: A Province - by - Province Study for China. ARC Centre of Excellence in Population Ageing Research (CEPAR). Working Paper. August 25, 2017.

[21] Zixi Li, Adam W. Shao, and Michael Sherris. The Impact of Systematic Trend and Uncertainty on Mortality and Disability in a Multistate Latent Factor Model for Transition Rates [J]. North American Actuarial Journal, 2017, 21, (4) : 594 - 610.

## A Study on Long Term Care Insurance Pricing Based on the Non - homogeneous Markov Model

Zhang Lin      Tang Wei

**Abstract :** This paper used the four - state time non - homogeneous Markov process to model the health transfer intensity of elderly people aged from 65 to 105 under the framework of Tweedie - GLM model, and also added factors such as age effects, time trends, and age - time interactions. The empirical results show that in addition to the common age effects, time trends and age - time interactions are also important factors in explaining the shift of health status. This paper also predicted life expectancy and expected time length of nursing needs based on actuarial theories, and created a long - term care insurance actuarial rating table for 2020. The results show that at all times and under the same age and initial health status conditions, females have an absolute survival advantage over males. Although males have a health advantage between 2002 and 2012, their disability time duration will last longer than females over time. But women's time duration of impairment has steadily surpassed that of men. As a combined result of these two situations, the premium rate of long - term care insurance for initially disabled males in 2020 is significantly higher than females.

**Key words :** non - homogeneous Markov; health transfer probability; disability; long - term care insurance