Insurance: Mathematics and Economics 52 (2013) 573–589

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: Application to long-term care insurance

Julien Tomas [1], Frédéric Planchet [*]

ISFA - Laboratoire SAF, France

HIGHLIGHTS

- We model the mortality of long-time care claimants having the same level of severeness.
- Adaptive local likelihood methods allow us to extract the information more pertinently.
- We vary the amount of smoothing in a location-dependent manner.
- We allow adjustments based on the reliability of the data.
- The potential uses of adaptive approaches suggest that they have much to offer.

ARTICLE INFO

Article history:
Received September 2012
Received in revised form
March 2013
Accepted 27 March 2013

Keywords:
Adaptive parameter choice
Local likelihood
p-splines
Long-term care insurance
Graduation

Mots clés:
Choix adaptif des paramètres
Vraisemblance locale
p-splines
Assurance dépendance
Graduation

ABSTRACT

We are interested in modeling the mortality of long-term care (LTC) claimants having the same level of severeness (heavy claimant). Practitioners often use empirical methods that rely heavily on expert opinions. We propose approaches not depending on an expert's advice. We analyze the mortality as a function of both the age of occurrence of the claim and the duration of the care. LTC claimants are marked by a relatively complex mortality pattern. Hence, rather than using parametric approaches or models with expert opinions, adaptive local likelihood methods allow us to extract the information from the data more pertinently. We characterize a locally adaptive smoothing pointwise method using the intersection of confidence intervals rule, as well as a global method using local bandwidth correction factors. The latter is an extension of the adaptive kernel method proposed by Gavin et al. (1995) to likelihood techniques. We vary the amount of smoothing in a location-dependent manner and allow adjustments based on the reliability of the data. Tests, and single indices summarizing the lifetime probability distribution are used to compare the graduated series obtained by adaptive local kernel-weighted log-likelihoods to *p*-spline and local likelihood models.

RÉSUMÉ

Nous nous intéressons à la construction de la loi de survie d'individus dépendants ayant le même niveau de sévérité (dépendance lourde). En pratique, les actuaires utilisent souvent des méthodes s'appuyant fortement sur l'opinion d'experts. Nous proposons des approches ne dépendant pas d'avis d'experts. La mortalité est analysée en fonction de l'âge à la survenance de la pathologie et l'ancienneté. La mortalité des dépendants est caractérisée par une structure relativement complexe. Plutôt que d'utiliser des approches paramétriques ou des modèles avec avis d'experts, les méthodes adaptatives de vraisemblance locale permettent d'extraire de façon pertinente l'information contenue dans les données en variant les paramètres du lissage selon l'âge à la survenance et l'ancienneté. Nous caractérisons une méthode ponctuelle de vraisemblance locale utilisant la règle de l'intersection des intervalles de confiance et un modèle global avec des facteurs d'ajustement local de la fenêtre d'observations. La dernière est une extension de la méthode adaptative à noyaux proposée par Gavin et al. (1995) aux techniques de

* Correspondence to: Institut de Science Financière et d'Assurances - Université Claude Bernard Lyon 1 - 50 Avenue Tony Garnier - 69366 Lyon, France.
E-mail addresses: julien.tomas@univ-lyon.fr (J. Tomas), frederic.planchet@univ-lyon1.fr, fred@planchet.net (F. Planchet).

[1] All computer programs are available on request.

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.insmatheco.2013.03.009

vraisemblance. Nous modifions le niveau de lissage en fonction de l'emplacement et nous permettons des ajustements de la fenêtre d'observations en fonction de la fiabilité des données. Des tests et marqueurs résumant les distributions de survie sont utilisés pour comparer les séries graduées obtenues par les méthodes adaptatives de vraisemblance locale aux modèles de *p*-splines et vraisemblance locale.

1. Introduction

In this article, we are interested in modeling the mortality of long-term care (LTC) claimants. LTC is a mix of social and health care provided on a daily basis, formally or informally, at home or in institutions, to people suffering from a loss of mobility and autonomy in their activity of daily living. Although loss of autonomy may occur at any age, its frequency rises with age. LTC insurance contracts are individual or collective, and they guarantee the payment of a fixed allowance, in the form of a monthly cash benefit, possibly proportional to the degree of dependency; see Kessler (2008) and Courbage and Roudaut (2011) for studies on the French LTC insurance market.

Most of the actuarial publications on this topic focus on the construction of models of projected benefits, see Deléglise et al. (2009), and modeling the life history of LTC patients using Markovian multistate models. Gauzère et al. (1999) model the progression of the pathologies. They use a non-parametric approach to derive smoothed estimates for the different transition intensities with a multistate model. This approach has the advantage of avoiding unreal assumptions, such as constant intensities resulting from the homogeneous Markov model. However, they assumes that the transition probabilities depend only on the age of the LTC patients. This assumption is inadequate when modeling heavy LTC claimant mortality. Czado and Rudolph (2002) assess the influence of factors like severeness of a pathology, gender, and type of care on the survival curve of the observed claims. Using the estimated hazard rates of the Cox (1972) proportional hazard model as transition intensities between levels of severity in a multiple state Markov model, they are able to fit a multiple state insurance model. In contrast to Cox's proportional hazard model, where the transition probabilities are calculated from the transition intensities, Helms et al. (2005) estimate the transition probabilities directly; they are then used to compute actuarial values of a given LTC plan and the required premiums. Recently, the risk coming from uncertainty in future demographical trends and the relevant impact on an LTC portfolio have been addressed by Levantesi and Menzietti (2012). Given a Markovian multistate mode, they define the benefits payable to the policyholder and the premiums payable to the annuity provider, when the insured is in a specific state. They propose then a stochastic model to assess the mortality and disability risk in life annuities with LTC benefits, and measure the risk evolution.

In health insurance, the effect of the age of the policyholder is often of unknown nonlinear form. In addition, neglecting the effect of calendar time of claims and district where the policyholder lives in modeling the claims process lead to biased fits, with corresponding consequences for risk premium calculation. Lang et al. (2002) present a space–time analysis of insurance data and allow one to explore temporal and spacial effects simultaneously with the impact of other covariates. They apply a semiparametric Bayesian approach for unified treatment of such effects within a joint model, developed in the context of generalized additive mixed models. Lang and Umlauf (2010) have extended a hierarchical version of regression models with structured additive predictor allowing nonlinear covariate terms in every level of the hierarchy. Their approach can deal simultaneously with nonlinear covariate effects and time trends, unit or cluster specific heterogeneity, spatial heterogeneity, and complex interactions between covariates of different type.

In contrast with the approaches exposed previously, we have no exogenous information about the LTC claimants, in terms of gender, place, or level of care. We observe only the aggregated exposition and number of deaths over two dimensions. These are the age of occurrence of the claim and the duration of the care. Here, LTC claimants belong only to one state of severeness (heavy claimants), and we are concerned with the construction of the survival distribution.

The pricing and reserving as well as the management of LTC portfolios are very sensitive to the choice of the mortality table adopted. In addition, the construction of such table is a difficult exercise for the following reasons.

i. The mortality law consists of a mixture of pathologies, and non-monotonic phenomena appear.
ii. French LTC portfolios are relatively small, and the estimation of crude death rates is very volatile.
iii. Because of the strong link between the age of occurrence of the claim and the related pathology, it is usual to construct a mortality table based on both age of occurrence of the claim, *which is an explanatory variable*, and duration of the care (or seniority), *which is the duration variable*. Hence, it is necessary to construct a mortality surface.
iv. Mortality rates decrease very rapidly with the duration of the care. In consequence, the first year is often difficult to integrate, disqualifying the usual parametric approaches.

Thus practitioners often use empirical methods that rely heavily on expert opinions. We therefore propose, in this article, methods not depending on expert advice, allowing us to extract the information from the data more pertinently. Unlike the problems presented in the literature above, this issue has not been addressed extensively.

We analyze the survival law as a function of both the age of occurrence of the claim and the duration of the care. The life table values computed are estimates of the true parameters, based on the finite amount of data available. The data examined should be regarded as a sample. Estimates based on the data will be subject to sampling errors, and the smaller the group is, the greater will be the relative random errors in the number of deaths and the less reliable will be the resulting estimates. However, we wish to smooth these quantities to reveal the characteristics of the mortality of LTC claimants, which we think to be relatively regular.

The article is organized as follows. Section 2 has still an introductory purpose. It makes precise the notation used in the following and introduces the ideas behind local likelihood fitting. Section 3 present the non-parametric estimators and covers model selection issues. The adaptive methods are introduced in Section 4. We characterize the intersection of confidence intervals rule and local bandwidth correction factors. Section 5 discusses the application on LTC claimants. Tests and single indices summarizing the probability lifetime distribution are used to compare the graduated series with those obtained from global non-parametric approaches, *p*-splines, and local likelihood, in Section 6. Finally, some remarks in Section 7 conclude the paper.