Technological Forecasting & Social Change 105 (2016) 89–93

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Latent heterogeneity effects in modelling individual hazards: A non-proportional approach

Renato Guseo *

Department of Statistical Sciences, University of Padua, Via C. Battisti 241, 35121 Padua, Italy

ARTICLE INFO

Article history:
Received 23 March 2015
Received in revised form 10 December 2015
Accepted 27 January 2016
Available online 15 February 2016

Keywords:
Heterogeneous diffusion models
Cox proportional hazards model
Bemmaor effects

ABSTRACT

This paper proposes an extension of individual hazard modelling through diffusion of innovation methodologies with the introduction of latent neighbouring effects in individual hazards. The proposal combines the Bemmaor's methodology for latent heterogeneous factors in the Cox proportional hazards model. The new model, the Bass–Bemmaor–Cox model, defines a non-proportional approach driven by a mixture of shifted Gompertz individual distributions where the time scale parameter governing growth dynamics depends upon observed covariates at the individual level. Conversely, individual propensity to the change of state included in the Gompertz sub-model is heterogeneous and latent within observed population. Based on experimental data referred to an ovarian pathology under treatment, an illustrative example is given. Further applications may be relevant for quantitative forecasting in marketing and technological diffusion contexts.

1. Introduction

Usually each individual event – a purchase, a shut-down, a death, a non-reversible change of state between two alternatives coded by zero and one – occurs in a precise moment that defines a duration from a conventional origin of the individual process.

Under convenient stationarity assumptions, the observed global duration in a population or system may be described through its empirical distribution over time. The construction of particular differential equations that determine the evolutionary shape of the counting process may be relevant, for interpretative and predictive aspects, within a theoretical framework. The basic idea in these models is that agents are not completely independent and their decisions (change of state) may be function of their interactions, usually latent. The logistic model by Verhulst (1838) is a pioneering reference in demography even if its main applications were conceived and are currently applied in biological and natural sciences contexts. A similar approach, typical of quantitative marketing modelling within diffusion of innovations methodology, is due to Bass (1969) with the introduction of distributed initializing seeding events through a monomolecular sub-model.

The introduction of *latent heterogeneity* of agents over time has defined an interesting perspective of research. Mainly, we highlight continuous and discrete directions, based on Bayesian mixtures, Bemmaor (1994), Bemmaor and Lee (2002) and Goswami and Karmeshu (2004), Guseo and Guidolin (2015), respectively.

There is a large body of literature that includes *observed heterogeneity* by micro-modelling the individual hazards with different perspectives. In particular, Chatterjee and Eliashberg (1990) consider heterogeneity in consumers' preference structure, modelling diffusion at the micro level, but application depends on specific information with related reliability and costs. Sinha and Chandrashekaran (1992) incorporate in a hazard model time-dependent observed covariates. This different perspective emphasizes an alternative approach which is strongly related to an individual description aiming at depicting the connection between an individual profile, characterized through covariates, and the particular observed decision between two mutually exclusive states. The effort is in modelling the time-dependent individual probability of such events.

In this field, a pioneering model is the Cox proportional hazards regression for survival data (Cox, 1972) where latent factors, which modulate interaction among agents and related decisions, are omitted, thus weakening interpretation and forecasting.

This paper focuses on the inclusion of *latent heterogeneity effects* in the Cox model through Bemmaor's approach. We call this new model the Bass–Bemmaor–Cox model in order to summarize the three main inspiring ideas by Bass, Bemmaor and Cox. In Section 2, we summarize the Cox model and then discuss the gamma-shifted Gompertz model by Bemmaor (1994) that includes the standard Bass one. We then propose a specific extension that combines individual and collective forces acting on a possible change of state of an agent. In Section 3, we illustrate

* Department of Statistical Sciences, University of Padua, via C. Battisti 241, 35123 Padua, Italy.
E-mail address: renato.guseo@unipd.it.

http://dx.doi.org/10.1016/j.techfore.2016.01.027
0040-1625/

an application on ovarian pathology data through a nonlinear regressive approach. Concluding remarks and comments are made in Section 4.

2. The Cox model and the gamma-shifted Gompertz distribution

Let us consider a random variable T that depicts a duration or a survival time. Its distribution $P(t) = Pr(T \le t)$ and the corresponding density $p(t) = dP(t)/dt$ may have a quite general shape. The complement of the distribution function, $S(t) = 1 - P(t)$, defines the so-called survival function. The hazard function, $h(t) = p(t)/S(t)$, describes the probability of an event (usually termed risk of the event) conditional on survival up to time t. Obviously, the shape of a hazard function depends upon the current distribution.

Cox (1972) considers a population of homogeneous independent individuals (agents) and states an individual hazard model $\lambda(t;z)$ that is proportional to a baseline $\lambda_0(t)$ through a function $H(z;\vartheta)$ that is a device to include covariates measured at individual level, $z = (x_0, x_1, x_2, x_3, \cdots)$, namely,

$$H(z;\vartheta) = \exp^{\vartheta_0 x_0 + \vartheta_1 x_1 + \vartheta_2 x_2 + \vartheta_3 x_3 + \cdots}. \quad (1)$$

Covariate x_0 may be, conventionally, a degenerate variable, e.g., $x_0 = 1$.

The proportional individual hazard proposed by Cox is therefore very simple

$$\lambda(t;z) = \lambda_0(t) \quad H(z;\vartheta). \quad (2)$$

Notice that the neutral level of $H(z;\vartheta)$ is 1, which corresponds to the non-significance of all covariates. Small levels of $H(z;\vartheta)$ depress the individual hazard, and conversely, high values of $H(z;\vartheta)$ augment the risk of event at study. The baseline hazard, $\lambda_0(t)$, is assumed common among independent agents in a population, and individual hazards are imposed to be strictly proportional to it defining a timeless condition over all eligible members of a population.

This aspect may be questionable because the particular choice $H(z;\vartheta)$ has no special theoretical meaning and is simply a pragmatical tool. Moreover, the major assumption is independence among individuals. In other words, the Cox model assumes a population of independent agents whose personal hazards are basically characterized by a function $H(z;\vartheta)$ of covariates z measured at individual level. There is no theoretical expression of interaction among individuals. The Cox model assumes that duration or survival time cannot be influenced by neighbouring or contagion (lato sensu) effects. This postulate is probably too strong in many situations where duration is partially dependent upon heterogeneous relationships among agents.

Complex Systems Analysis and related modelling tools are mainly grounded on the opposite assumption giving rise to Cellular Automata and Network Automata models, Boccara (2004). The key idea is that agents in a population are certainly characterized by heterogeneous individual levels of some factors, but only some of them, that are crucial for a change of state, may be hiddenly modified over time by the influence of social networks. Interaction, which is mostly generated within thematic network relationships among agents, is due to hierarchical influences of special 'hubs' or local 'social' forces. In particular, Barabasi (2012) discusses the role of preferential attachment in network science; Newman et al. (2011) and Liu et al. (2013) examine the characterizations and dynamics of observable networks; Krivitsky and Handcock (2014) and Handcock and Gile (2010) consider the statistical inferential aspects in modelling networks with sampled data, and Guseo and Guidolin (2009) model and identify a latent network of relationships in determining a dynamic market potential in a diffusion of innovations context.

Diffusion of innovation methodology and related aggregate models are obtained by expressing a duration framework through special differential equations that depict theoretically the collective behaviour of growing systems. To this end, see, in particular, Bass (1969) for a basic model, and Bass et al. (1994) for a quite general solution that includes controlling covariates that modify the dynamics of the entire system. Further advances consider: multi-cyclic situations, Robertson et al. (2007) and Guseo and Guidolin (2015), among others; non-constant carrying capacity or dynamic market potentials, Mahajan and Peterson (1978, 1985) and Guseo and Guidolin (2009); competition between products or brands in a marketplace, Savin and Terwiesch (2005), sequential entries in competing contexts, Guseo and Mortarino (2012), and separable word-of-mouth estimation of competing products, Guseo and Mortarino (2014).

Previous models assume local homogeneity of the involved few subgroups in a population. However, heterogeneity and, in particular, latent heterogeneity may strongly affect agents' behaviour and their dynamics in a duration framework.

Bemmaor (1994) and Bemmaor and Lee (2002) suggest a simple unimodal methodology for explaining local changes in the parameter estimates of the basic Bass model due to underlying latent heterogeneity of the population. They consider, in particular, a random variation in individual propensities to buy assumed to follow a gamma distribution and describe the timing of the first purchase through a shifted Gompertz density. The aggregate diffusion process results in a composition of these two densities that we call Bemmaor's approach.

More formally, the individual-level model for the adoption time description is specified with a shifted Gompertz distribution, i.e.,

$$F_{SG}(t) = \left(1 - e^{-bt}\right) e^{-\eta e^{-bt}}, \quad t > 0, \quad \eta, b > 0. \quad (3)$$

For simplicity, parameter b is assumed fixed. Small values of η denote low individual mean adoption times defining, therefore, high propensity to buy or to a change of state between two alternatives.

In the sequel we modify Bemmaor's approach by introducing covariates z. Notice that parameter b in Eq. (3) has the dimension of time and, therefore, we may model it through the function $H(z;\vartheta)$ defined in Eq. (1) in order to take into account individual dynamics, by assuming,

Table 1
Ovarian data: *np*, patient label; *t*, elapsed time. Its meaning depends upon the indicator variable *y*. If $y = 0$, patient is still alive, and *t* denotes current elapsed time since initial diagnosis. If $y = 1$, patient is not alive, and *t* represents a duration. Individual factors: x_1, geographic area; x_2, treatment group; x_3, ultrasound response; and x_4, patient age.

np	t	y	x_4	x_1	x_2	x_3
1	59	0	72.3315	2	1	1
2	115	0	74.4932	2	1	1
3	156	0	66.4658	2	1	2
4	421	1	53.3644	2	2	1
5	431	0	50.3397	2	1	1
6	448	1	56.4301	1	1	2
7	464	0	56.937	2	2	2
8	475	0	59.8548	2	2	2
9	477	1	64.1753	2	1	1
10	563	0	55.1781	1	2	2
11	638	0	56.7562	1	1	2
12	744	1	50.1096	1	2	1
13	769	1	59.6301	2	2	2
14	770	1	57.0521	2	2	1
15	803	1	39.2712	1	1	1
16	855	1	43.1233	1	1	2
17	1040	1	38.8932	2	1	2
18	1106	1	44.6	1	1	1
19	1129	1	53.9068	1	2	1
20	1206	1	44.2055	2	2	1
21	1227	1	59.589	1	2	2
22	268	0	74.5041	2	1	2
23	329	0	43.137	2	1	1
24	353	0	63.2192	1	2	2
25	365	0	64.4247	2	2	1
26	377	1	58.3096	1	2	1