

Article

Estimation–Calibration of Continuous-Time Non-Homogeneous Markov Chains with Finite State Space

Manuel L. Esquível ^{1,*}, Nadezhda P. Krasii ² and Gracinda R. Guerreiro ¹

¹ Department of Mathematics, NOVA FCT, and NOVA Math, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Monte de Caparica, Portugal; grg@fct.unl.pt

² Department of Higher Mathematics, Don State Technical University, Gagarin Square 1, Rostov-on-Don 344000, Russia; spu-46.3@donstu.ru

* Correspondence: mle@fct.unl.pt; Tel.: +351-212-948-388

Abstract: We propose a method for fitting transition intensities to a sufficiently large set of trajectories of a continuous-time non-homogeneous Markov chain with a finite state space. Starting with simulated data computed with Gompertz–Makeham transition intensities, we apply the proposed method to fit piecewise linear intensities and then compare the transition probabilities corresponding to both the Gompertz–Makeham transition intensities and the fitted piecewise linear intensities; the main comparison result is that the order of magnitude of the average fitting error per unit time—chosen as a year—is always less than 1%, thus validating the methodology proposed.

Keywords: Markov chains; non homogeneous; continuous time; regime switching processes; estimation; calibration; health insurance; long-term care

MSC: 60J27; 65C40; 60J28; 62M05



Citation: Esquível, M.L.; Krasii, N.P.; Guerreiro, G.R. Estimation–Calibration of Continuous-Time Non-Homogeneous Markov Chains with Finite State Space. *Mathematics* **2024**, *12*, 668. <https://doi.org/10.3390/math12050668>

Academic Editors: Panagiotis-Christos Vassiliou and Andreas C. Georgiou

Received: 2 January 2024

Revised: 20 February 2024

Accepted: 21 February 2024

Published: 24 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction with a Literature Review

This study follows on from a previous article (see [1]) in which we developed a way to calibrate a Markov chain model in continuous time using data obtained from Portuguese National Network of Continuing Care. The calibration methodology used in that work, although very effective, is not completely satisfactory as it rests on a series of ad hoc processes with reduced guarantees of reproducibility and robustness.

In the present work, we intend to develop simpler and more robust means of estimating and calibrating intensities for non-homogeneous continuous-time Markov chains (see [2] for a recent introduction to these processes and their applications). For this purpose, we first develop the two following subjects. The first subject deals with Markov chain regime switching achieved by considering an abrupt change in the intensities, for instance, having intensities with jumps. The second subject complements the first one if we suppose that we replace regular intensities by irregular ones—like piecewise linear—in principle with more easily estimable parameters; we study the effect on the transition probabilities of a replacement of the original intensities by sufficiently close alternative intensities. These two different streams of ideas are connected not only to one another but also to the estimation–calibration techniques to be studied.

We now present a review of the literature, mainly covering the subject of estimation and calibration of continuous-time non-homogeneous Markov chains with finite state space relevant for health insurance and long-term care (LTC), existing results for the Kolmogorov ordinary differential equations, as well as works where one can find some similarities between non-homogeneous Markov chains and semi-Markov jump linear systems.

A consecrated approach in the study of continuous-time Markov chains for applications, namely, in the multiple state models—the transition intensity approach (see [3] p. 126 or [4] p. 189)—consists of giving the intensities, solving the Kolmogorov ODE and using

the transition probabilities obtained for computations. The intensities should be estimated from the data. This is the approach that we assume in this work.

The statistics of homogeneous Markov chains has already received several very thorough analyses. A very well organised and complete one is provided in Billingsley's monograph [5] that treats, in the first part, the discrete-time homogeneous Markov chains and in the second part the continuous-time chains by resorting to the canonical embedded process. A companion reference is article [6] that provides a very complete set of references on the subject until 1961. In order to obtain consistency and asymptotic normality results for the maximum likelihood estimators, the author assumes, as usual, stringent regularity assumptions in particular on the intensities.

The statistics of Markov chain models for multiple state models is usually performed under simplifying assumptions on the model. For instance, in [7], the intensities are supposed to be constant in selected time intervals, and observations are chosen for which the exact age belongs to a given selected time interval. Another set of simplifying assumptions is proposed in [8] (pp. 126–128); at first, the transition functions are approximated in a one-year period interval by a one-sided Stieltjes interval. and then, using these approximations the transition intensities, with adequate analytical properties, are obtained as a result of a minimisation of a sum of squares objective function. The method proposed in [4] (pp. 147–169) has also two steps; in the first step, the transitions intensities are supposed to be constant in one-year period intervals and are estimated with a maximum likelihood approach. Subsequently, there is a second step of denominated graduation—a method generally described in [9]—that fits parameters of exponential functional intensity using generalised linear models. The method is applied to real data, and it becomes clear that several adequate particular ad hoc assumptions in the method are inevitable in order to deal with specific properties of the data. The simplifying assumption of transition intensities constant in each one-year period is also taken in [10] (pp. 683–690), where a detailed treatment of an example is also presented; in a commentary, the authors also refer the need of a graduated procedure to obtain the final intensities.

The excellent review work [11] illustrates the manner in which multiple-state Markov and semi-Markov models can be used for the actuarial modelling of health insurance policies. The bivariate character of the Markov process naturally associated with a semi-Markov model is useful whenever the durational effects are not negligible but in contrast is technically much more difficult to handle than the univariate Markov process. Considering discrete-time semi-Markov processes, the authors in [12] study semi-Markov jump linear systems—which is a hybrid dynamical system that consists of a family of subsystem modes and a semi-Markov process that orchestrates switching between them—with bounded sojourn times, in order to provide sufficient criteria for the stability and stabilisation problems with respect to a specified approximation error. The companion work [13] enlarges the previous model by considering delay, and by means of a novel Lyapunov–Krasovskii functional and using the probability structure of semi-Markov switching signal, the sufficient stability conditions for the considered systems are presented in terms of a set of linear matrix inequalities and a proper semi-Markov switching condition. It now becomes clear that a natural extension of our work would be to consider semi-Markov models instead of Markov chains.

The estimation–calibration methodology we propose in this work is applied to continuous piecewise linear intensities. Since in health insurance and long-term-care multiple-state models, the intensities are usually of Gompertz–Makeham type (see [4] pp. 21, 24, 101), we previously showed that the distance between two transition probability matrices—in the sense of some matrix norm—is bounded by the same distance between the correspondent intensity matrices, thus showing that the Gompertz–Makeham functional form for the intensities is not really necessary.

We now refer to some works with contributions to the topic of estimation calibration of multiple state models, also called multi-state models. The work [14] proposes a review of multiple-state models via continuous-time Markov chains, signalling the usual approach for the non-homogeneous case of considering piecewise constant intensities. It is at first reading, a most interesting review paper with applications to real data comparing different model analyses. Ref. [15] deals with a nonparametric approach to statistical inference in non-homogeneous Markov processes based on counting processes for transition intensities—namely using the so-called Nelson–Aalen estimator or the kernel smoothing estimator of Ramlaou–Ahnsen—presenting a case study using this methodology. Ref. [16] can be seen as a continuation of the previously referred to work. Besides reviewing methods for non-parametric estimation of transition probabilities, the authors study the case where semi-parametric Cox type regression models are specified for the transition intensities whenever there is specification of the development of the time-dependent covariates. An illustration of the methods with data from a randomised clinical trial in patients with liver cirrhosis is also presented. Ref. [17] is an ancillary reference for graduating the transition intensities in a multiple-state model for permanent health insurance applications based on generalised linear models—with a random component based on independent Poisson response variables—in the case that the intensities are supposed to depend on some secondary variables. The work in [18] follows the preceding paper in the main intent of proposing a graduation method for the transition intensities of a non-homogeneous continuous-time Markov chain model. In the work [19], a comparison between a discrete-time and continuous-time homogeneous Markov chain models is presented in order to assess the effect of unevenly spaced observations. Since the authors want to incorporate covariates in the model, this study also deals with a series of multinomial logit regressions for the discrete-time model and proportional hazard regressions for the covariates through transition intensity functions for the continuous-time model. Ref. [20] is a simplified multiple-state model that develops a generic estimation method for calculating the transition probabilities in a one-year multiple-state model based on disability prevalence rates; multiple logistic regression models are employed to estimate disability prevalence rates and the one-year recovery rates. In doing so, the authors assume three conditions of the ratio between the mortality rate of inactive and active people—and several other conditions used in the literature—that allow the necessary computations in the case treated which concerns cross-sectional data measuring the disabled status of an individual at one point in time. The work [21] introduces a semi-parametric model that employs a logit function to capture the treatment intensities across two groups, aiming to estimate transition intensity rates within the framework of an illness–death model. Parameter estimation is conducted through an EM algorithm coupled with profile likelihood. Simulation studies presented in the text indicate that the proposed method is straightforward to apply and produces results comparable to those of the parametric model. The study [22] examines the impact of part-time and full-time employment on health by employing a Markov three state model—using piecewise constant forces, where the transition intensities are graduated using generalised linear models and assumed, at the start, to be equal per age level—and generalised linear models to refine the initial raw rates. Integration of the corresponding Chapman–Kolmogorov equations allow us to derive a comprehensive solution. As an application of the model, the effectiveness of a partial early-retirement incentive in the Netherlands is evaluated. The refined rates obtained indicate that working part-time does not necessarily correlate with improved health among the elderly.

In the present work, we also establish a result with regime-switching Markov chains exploring the possibility of having, in the whole time period under study, intensities with several different functional forms (linear, exponential, etc.) in different subintervals of the whole time period. Our study of ordinary differential Equations (ODE) with regimes that began with the work [23] and was exploited in [24] is based on general results of existence and uniqueness of solutions of ODE—due to Caratheodory and Wintner for existence and Osgood for uniqueness, among others—with a non-regular second member. In the case of

a second member of the non-regular ODE, the regimes appear clearly in the solutions of the Kolmogorov equations due, for example, to possible discontinuities in the entries of the intensity matrix.

The general theory of existence of solutions of Kolmogorov equations is exhaustively treated in the works of Feinberg, Shiryaev and Mandala (see [25–27]). Their powerful results apply to jump processes with values in a general Borel space and so are also transferable to non-homogeneous Markov chains with a finite state space. Since we are dealing with this finite state space—due to our interest in health insurance and long-term-care multiple-state models—we chose to present a more simple approach that requires only classical existence theorems for ordinary differential equations, namely Caratheodory’s existence theorem and Osgood’s uniqueness theorem.

We now succinctly describe the contents of this work.

- In Section 2, we develop the subject of regime-switching Markov chains. The results obtained can be applied to the consideration of discontinuous intensity matrices.
- In Section 3, we deal with the approximation of matrices of transition probabilities given an approximation of the correspondent matrices of intensities.
- Sections 4–6 detail, with an example, the methodology for estimation–calibration proposed and present an analysis of the results obtained.
- In Section 7, we provide a discussion of the results obtained in the example treated, and in Section 8, we summarise all the results obtained in this work.

There are three main contributions of this work. The first is the proposal of a method to estimate the parameters of a set of transition intensities from ideal observed data. The second is a result on regime-switching Markov chains that establishes the possibility of considering transition intensities made up of different sorts of functional forms, with each one of the functional forms depending on different sets of parameters. Finally, the third contribution is a result that quantifies the norm of the difference of two probability transition matrices in terms of the norm of the corresponding matrices of transition intensities; this last result justifies the choice of arbitrary functional forms for the transition intensities in ways more adequate for parameter estimation.

2. Regime Switching Markov Chains

In this section, we develop the formalism of regime switching for Markov chains, in which the transition probabilities are derived from intensities that, at a certain point in time, can change either in functional form or in the parameters. The consideration of discontinuous piecewise linear intensities suggests the study of Markov chains in continuous time with regimes. Let us state some preliminary notations and results for context purposes (see [28]). Firstly, we recall the definition of an intensity matrix $Q(t, \theta)$.

Definition 1. Let $\mathcal{L}(\mathbb{R}^{d \times d})$ be the space of $d \times d$ square matrices with coefficients in \mathbb{R} . A function $Q : [0, +\infty[\rightarrow \mathcal{L}(\mathbb{R}^{d \times d})$ denoted by

$$Q(t, \theta) = [\mu_{ij}^\theta(t)]_{i,j=1,\dots,d},$$

with $\theta \in \Theta \subset \mathbb{R}^p$ a parameter is a transition intensity matrix if, for almost all $t \geq 0$, it verifies

- $\forall i = 1, \dots, d, t \geq 0, \mu_{ii}^\theta(t) \leq 0;$
- $\forall i = 1, \dots, d, \forall j = 1, \dots, d, t \geq 0, i \neq j \Rightarrow \mu_{ij}^\theta(t) \geq 0;$
- $\forall i = 1, \dots, d, t \geq 0, \sum_{j=1,\dots,d} \mu_{ij}^\theta(t) = 0.$

Secondly, we recall the Kolmogorov ordinary differential Equations (ODE) for non-homogeneous continuous-time Markov chains. These equations, upon integration, give the matrix of transition probabilities $P(x, t, \theta)$ as a function of the matrix of intensities $Q(t, \theta)$, where $\theta \in \Theta$ is a parameter. The forward Kolmogorov ODE can be represented in the following form:

$$\begin{cases} P'_t(x, t, \theta) = P(x, t, \theta)Q(t, \theta) \\ P(x, x) = I \end{cases}, \quad (1)$$

or, in integrated form, by

$$P(x, t, \theta) = I + \int_{[x, t]} P(x, s, \theta)Q(s, \theta)ds. \quad (2)$$

Finally, let us now deal with regime-switching Markov chains.

The general motivation for the study of regime switching in ODE can be seen in [23,24]. Let us elaborate on the motivation to study regime-switching Markov chains. Suppose that we have two continuous-time Markov chains, having the same state space, with intensities of different functional forms—such as piecewise constant or affine of the Gompertz-Makeham type, etc.—depending on two sets of parameters, say Θ_1 and Θ_2 , respectively, defined in two contiguous-time intervals, say $[0, t_1]$ and $[t_1, t_2]$. The result we prove shows that there exists a well-defined Markov chain in the interval $[0, t_2]$ with intensities depending on a set of parameters $\Theta := \Theta_1 \cup \Theta_2$ such that the transition probabilities of this Markov chain—obtained by the solution of the Kolmogorov equations—coincide with the transition probabilities of the first initial Markov chain in $[0, t_1]$ and also coincide with the transition probabilities of the second initial Markov chain in $[t_1, t_2]$. This result will grant us a greater latitude in the choice of the functional forms of the intensities for estimation purposes since it will be possible to partition a time interval of interest in two or more disjoint intervals and to have intensities, in each one of the intervals, possibly of different functional forms and different sets of parameters. In Theorem 4, we consider extended solutions of an ODE in the sense of Carathéodory. For that, following [29] (pp. 41–44), we consider the definition of an extended solution of a differential equation.

Definition 2 (Extended solution of an ODE). For $f(t, y) : I \times \mathcal{D} \rightarrow \mathbb{R}^{d \times d}$ a non necessarily continuous function, with $I \subset [0, +\infty[$ and $\mathcal{D} \subset \mathbb{R}^{d \times d}$ and a differential equation given by

$$Y'(t) = f(t, Y(t)), \quad Y(0) = y_0 \in \mathbb{R}^{d \times d}, \quad (3)$$

or in the equivalent integral form, with the appropriate Lebesgue measure du in \mathbb{R} ,

$$Y(t) = y_0 + \int_0^t f(u, Y(u))du, \quad (4)$$

an extended solution $Y(t)$ of the ODE in Formula (3) is an absolutely continuous function $Y(t)$, such that $f(t, Y(t)) \in \mathcal{D}$ for $t \in I$ and Formula (3)—or equivalently, Formula (4)—is verified for all $t \in I$ almost everywhere (a.e), that is, possibly with the exception of a set of null Lebesgue measure in $[0, +\infty[$.

We now recall Carathéodory's existence theorem—see [29] (p. 43) for the unidimensional result and [30] (pp. 28–29) for the multidimensional result, with a proof via Schauder's fixed point theorem—in the context of the model we are studying, a theorem that ensures the existence of an extended solution under general conditions.

Theorem 1 (Caratheodory's existence theorem). Suppose that $f(t, \mathbf{y}) : I \times \mathcal{D} \rightarrow \mathbb{R}^{d \times d}$, with $I = [0, u[$ an open set of $[0, +\infty[$ and \mathcal{D} an open set of $\mathbb{R}^{d \times d}$, verifies that

- (i) $f(t, \mathbf{y})$ is measurable in the variable t , for fixed \mathbf{y} , and continuous in the variable \mathbf{y} , for fixed t , for $(t, \mathbf{y}) \in I \times \mathcal{D}$.
- (ii) For each compact set $K \Subset \mathcal{D}$ and $T > 0$, there exists a Lebesgue integrable function $\lambda(t)$, such that $\|f(t, \mathbf{y})\| \leq \lambda(t)$ for $(t, \mathbf{y}) \in [0, T] \times K$.

Then, for every $(t_0, \mathbf{y}_0) \in I \times \mathcal{D}$ such that $\mathbf{Y}(t_0) = \mathbf{y}_0$, that is, a given initial condition of equation in Formula (3), there exists an extended solution according to Definition 2, defined in a neighbourhood of (t_0, \mathbf{y}_0) .

Despite the fact that Theorem 1 guaranties the local existence of an extended solution, it is always possible to consider a maximal extension of this solution, possibly, to a larger time interval (see [30] pp. 29–30).

Theorem 2 (Maximal time interval for existence). With the notations and under the hypothesis of Theorem 1, any existing solution \mathbf{Y} admits a continuation $\tilde{\mathbf{Y}}$ to a maximal time interval of existence, let it be $[a, b]$ such that, $\partial\mathcal{D}$ being the boundary of \mathcal{D} :

$$\lim_{t \rightarrow a} \tilde{\mathbf{Y}}(t) \in \partial\mathcal{D} \text{ and } \lim_{t \rightarrow b} \tilde{\mathbf{Y}}(t) \in \partial\mathcal{D}.$$

Remark 1 (Applying Caratheodory theorem to the Kolmogorov ODE). Kolmogorov ODE for continuous time Markov chains, in Formula (1), falls under this formalism in the following way:

$$f(t, \mathbf{y}, \boldsymbol{\theta}) = \mathbf{Q}(t, \boldsymbol{\theta}) \cdot \mathbf{y} \quad (5)$$

which is essentially the equation in Formula (3) with the possibility of dependence on a parameter $\boldsymbol{\theta} \in \Theta$. Consider a matrix norm $\|\cdot\|$ in the sense of [31] (p. 340), that is, a submultiplicative norm—such as the l^1 norm and l^2 norm, also known as the Frobenius norm—and observe that since

$$\|f(t, \mathbf{y}, \boldsymbol{\theta})\| = \|\mathbf{Q}(t, \boldsymbol{\theta}) \cdot \mathbf{y}\| \leq \|\mathbf{Q}(t, \boldsymbol{\theta})\| \|\mathbf{y}\|, \quad (6)$$

and since any norm of a probability intensity matrix is bounded, we can apply Caratheodory's theorem to the Kolmogorov ODE under the condition that there exists a Lebesgue integrable function $\lambda(t)$ such that

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{Q}(t, \boldsymbol{\theta})\| \leq \lambda(t), \quad (7)$$

for all $t \in [0, T]$. We will see in Remark 3 that the condition in Formula (7) is also sufficient to ensure the unicity of the extended solutions.

Remark 2 (Existence of extended solutions alternative proof). We could also quote Wintner's theorem, referred to in [32], a theorem that states that if $\|f(t, \mathbf{y})\| \leq N(t)L(\|\mathbf{y}\|)$ with N and L piecewise continuous, positive and L non-decreasing, such that for some $c > 0$

$$\int_c^{+\infty} \frac{1}{L(s)} ds = +\infty,$$

then the ODE in Formula (3) has a solution for a given initial condition. We observe that the quoted theorem is valid, under the assumption that $f(t, \mathbf{y})$ is continuous with the possible exception of points of a null Lebesgue set of the time variable, by considering extended solutions instead of usual solutions, which have a continuous derivative, and as $L(t)$ and $N(t)$ both satisfy the hypotheses of Wintner's theorem for parameters of the intensity functions and taking each, two (or several) distinct values in two (or several) complementary intervals of the time domain.

We now quote an unicity result—from [30] p. 30—applicable whenever there is existence of an extended solution in the Caratheodory sense.

Theorem 3 (Unicity of extended solutions). *Suppose that $f(t, \mathbf{y}) : I \times \mathcal{D} \rightarrow \mathbb{R}^{d \times d}$, with $I = [0, u[$ an open set of $[0, +\infty[$ and \mathcal{D} an open set of $\mathbb{R}^{d \times d}$, verifies the conditions in Theorem 1, and moreover, that for each compact set $K \subseteq \mathcal{D}$ and $T > 0$, there exists a Lebesgue integrable function $\lambda_K(t)$, such that*

$$\|f(t, \mathbf{y}_1) - f(t, \mathbf{y}_2)\| \leq \lambda_K(t) \|\mathbf{y}_1 - \mathbf{y}_2\|, \quad (8)$$

for $(t, \mathbf{y}_1), (t, \mathbf{y}_2) \in [0, T] \times K$. Then, for every $(t_0, \mathbf{y}_0) \in I \times \mathcal{D}$ such that $\mathbf{Y}(t_0) = \mathbf{y}_0$, that is, a given initial condition of equation in Formula (3), there exists an unique extended solution \mathbf{Y} according to Definition 2, defined in a neighbourhood of (t_0, \mathbf{y}_0) . The domain of definition of \mathbf{Y} is open, and \mathbf{Y} is continuous in this domain.

Remark 3 (Applying the unicity result to the Kolmogorov ODE). *With the interpretation given in Formula (5) and if the norm is a matrix norm, similarly to what we had in Formula (6), we now have that a sufficient condition for the unicity of the extended solutions of Kolmogorov ODE is for each compact set $K \subseteq \mathcal{D}$ and $T > 0$, there exists a Lebesgue integrable function $\lambda_K(t)$ such that $\|Q(t, \theta)\| \leq \lambda_K(t)$, thus implying that*

$$\begin{aligned} \|f(t, \mathbf{y}_1, \theta) - f(t, \mathbf{y}_2, \theta)\| &= \|Q(t, \theta) \cdot (\mathbf{y}_1 - \mathbf{y}_2)\| \leq \|Q(t, \theta)\| \|\mathbf{y}_1 - \mathbf{y}_2\| \leq \\ &\leq \lambda_K(t) \|\mathbf{y}_1 - \mathbf{y}_2\|, \end{aligned} \quad (9)$$

which is the hypothesis bound in Formula (8) for Theorem 3.

Remark 4 (On the unicity of the extended solutions). *Either directly using Theorem 18.4.13 in [33] (p. 337) or using Osgood's uniqueness theorem—as presented for instance, in [34] (p. 58) or in [35] (pp. 149–151)—we may also conclude that the extended solution, that we know to exist, is unique, in the sense that two solutions may only differ on a set of Lebesgue measures equal to zero.*

Remark 5 (On the numerical computation of extended solutions). *We observe that these existence and uniqueness results are essential for a numerical integration of the ODE, but that no result on numerical convergence is implied—in the existence and uniqueness results above—for the regime switching ODE with discontinuous coefficients. Nevertheless, the Lipschitz condition with respect to the \mathbf{y} variable—such as the one in Formula (8)—is sufficient for the convergence of the Euler method (see [36] p. 74).*

The next theorem is a simple example of a regime-switching result for continuous-time Markov chains. The extension of this result to more than two regimes is straightforward. We consider the Kolmogorov ODE in a time interval $[0, T]$.

Theorem 4 (Regime switching continuous-time Markov chains). *Let $\|\cdot\|$ denote a matrix norm, let Θ denote a parameter set and let $Q_1(t, \theta)$ defined for $t \in [0, t_1]$ and $Q_2(t, \theta)$ defined for $t \in [t_1, T]$ be two intensity matrices such that for $\lambda(t)$, an integrable function defined in $[0, T]$, we have for $t \in [0, T]$*

$$\max \left(\sup_{\theta \in \Theta} \|Q_1(t, \theta)\|, \sup_{\theta \in \Theta} \|Q_2(t, \theta)\| \right) \leq \lambda(t). \quad (10)$$

Then, there exists $\tilde{P}(t, \theta)$ such that

1. In $[0, t_1]$, we have that $\tilde{P} \equiv P_1$, a.e. in t , where P_1 is a solution of the Cauchy problem $(P_1)'_t(t, \theta) = P_1 Q_1(t, \theta)$ with the usual initial conditions;
2. In $[t_1, T]$, we have that $\tilde{P} \equiv P_2$ a.e. in t , where P_2 is a solution of the Cauchy problem $(P_2)'_t(t, \theta) = P_2 Q_2(t, \theta)$ with the initial conditions given by $P_1(t_1, \theta)$;
3. \tilde{P} is a transition probability matrix.

Proof. Let $Q_1(t, \theta) = [\mu_{ij}^{1,\theta}(t)]_{i,j=1,\dots,d}$ and $Q_2(t, \theta) = [\mu_{ij}^{2,\theta}(t)]_{i,j=1,\dots,d}$. If we define

$$Q(t, \theta) = [\mu_{ij}^{\theta}(t)]_{i,j=1,\dots,d} := Q_1(t, \theta)\mathbb{I}_{[0,t_1]}(t) + Q_2(t, \theta)\mathbb{I}_{]t_1,T]}(t), \quad (11)$$

we will have that:

$$\mu_{ij}^{\theta}(t) = \mu_{ij}^{1,\theta}(t)\mathbb{I}_{[0,t_1]}(t) + \mu_{ij}^{2,\theta}(t)\mathbb{I}_{]t_1,T]}(t),$$

and we can, immediately, verify that $Q(t, \theta)$ is an intensity matrix on $[0, T]$ according to Definition 1. Moreover, since by Formula (10) and the definition in Formula (11), we have that

$$\|Q(t, \theta)\| \leq \lambda(t),$$

we can let $\tilde{P}(t, \theta)$ be the unique solution of the Kolmogorov equation $\tilde{P}'_t = \tilde{P}Q$ on $[0, T]$ with the usual conditions. It is then clear that $\tilde{P}(t, \theta)$ is a transition probability matrix. Furthermore, if we define

$$\hat{P}(t, \theta) := P_1(t, \theta)\mathbb{I}_{]0,t_1[}(t) + P_2(t, \theta)\mathbb{I}_{]t_1,T[}(t)$$

and we will then have, using the hypothesis that

$$\begin{aligned} \frac{\partial \hat{P}(t, \theta)}{\partial t} &\stackrel{\text{a.e.}}{=} \frac{\partial \hat{P}_1(t, \theta)}{\partial t}\mathbb{I}_{]0,t_1[}(t) + \frac{\partial \hat{P}_2(t, \theta)}{\partial t}\mathbb{I}_{]t_1,T[}(t) \stackrel{\text{a.e.}}{=} \\ &\stackrel{\text{a.e.}}{=} P_1 Q_1(t, \theta)\mathbb{I}_{]0,t_1[}(t) + P_2 Q_2(t, \theta)\mathbb{I}_{]t_1,T[}(t) \stackrel{\text{a.e.}}{=} \\ &\stackrel{\text{a.e.}}{=} \left(P_1(t, \theta)\mathbb{I}_{]0,t_1[}(t) + P_2(t, \theta)\mathbb{I}_{]t_1,T[}(t) \right) \left(Q_1(t, \theta)\mathbb{I}_{[0,t_1]}(t) + Q_2(t, \theta)\mathbb{I}_{]t_1,T]}(t) \right) \stackrel{\text{a.e.}}{=} \\ &\stackrel{\text{a.e.}}{=} \hat{P}(t, \theta)Q(t, \theta), \end{aligned}$$

which shows that $\hat{P}(t, \theta) \equiv \tilde{P}(t, \theta)$ a.e. in t and, as a consequence, that $\tilde{P}(t, \theta) \equiv \hat{P}(t, \theta) \equiv P_1$ in $[0, t_1]$ a.e. and $\tilde{P}(t, \theta) \equiv \hat{P}(t, \theta) \equiv P_2$ in $[t_1, T]$ a.e. \square

We present in Figure 1 graphical representations of transition probabilities obtained by numerical integration of Kolmogorov equations with discontinuous piecewise linear intensities for a four-state Markov chain with intensity matrix given in Formula (16).

State one is the healthy state, state four is an absorbing state corresponding to death and states two and three are intermediate dependent states. This representation, aside from being an illustration of a regime switching Markov chain, also illustrates the possible extreme effects of a regime switching of discontinuous-intensity matrix entries. The subject of Markov chains with regimes has, as can easily be observed, an interest independent of the objective that motivates us to study it. However, in the context of the present work, it can be a way to justify a more efficient and robust parameter estimation (or calibration) process by an adequate choice of functional forms for the intensities.

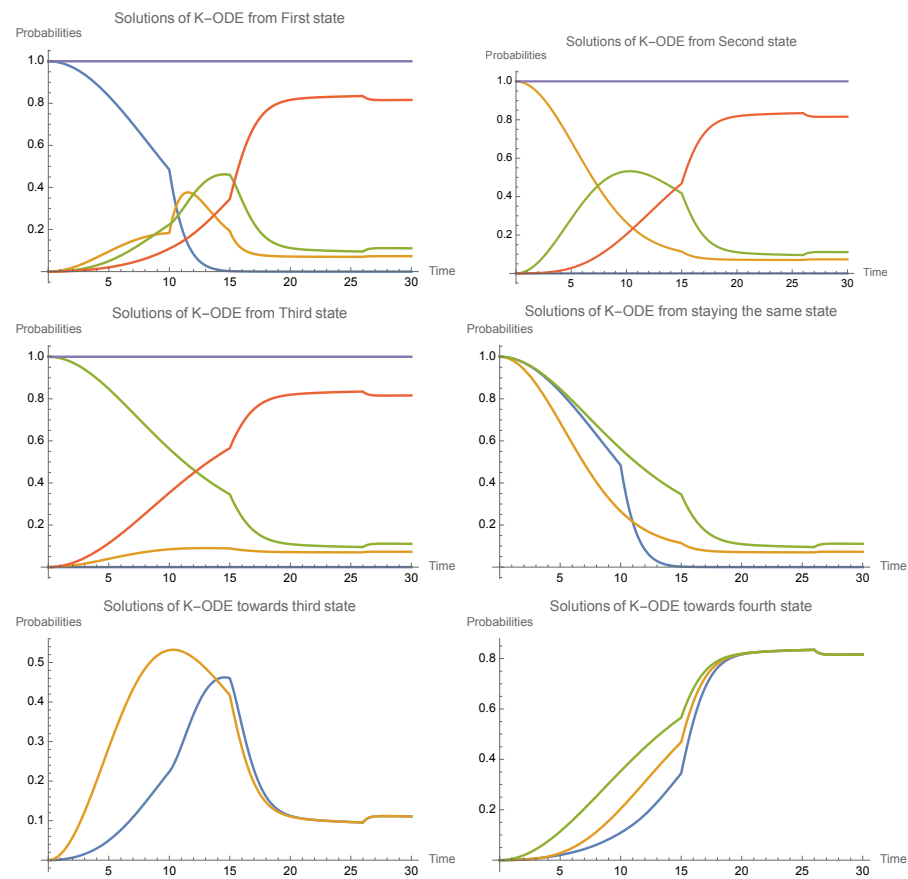


Figure 1. Solutions of Kolmogorov ODE for discontinuous linear intensities. p_{ij} : $j = 1$ (blue); $j = 2$ (orange); $j = 3$ (green); $j = 4$ (red). $\sum_{j=1}^4 p_{ij}$ (purple). Lower left p_{i3} : $i = 1$ (blue); $i = 2$ (orange); Lower right p_{i4} : $i = 1$ (blue); $i = 2$ (orange); $i = 3$ (green).

3. On the Approximation of Intensities and Corresponding Transition Probabilities

One way to simplify the estimation of intensities of Markov chain models in continuous time—relevant for health insurance and long-term-care models—is to replace the usual Gompertz–Makeham type intensities—containing exponential and linear terms and therefore being difficult to estimate—by continuous piecewise linear intensities. In this sense, it is important to have a result that controls the distance between two matrices of transition probabilities resulting from the integration of the Kolmogorov equations for the correspondent two matrices of intensities.

It is known (see, for example, [37] p. 264 and [1]) that we can represent the transition probabilities in the Hostinsky form:

$$P(x, t, \theta) = I + \sum_{n=1}^{+\infty} \int_{[x,t]} \int_{[t_1,t]} \cdots \int_{[t_{n-1},t]} Q(t_1, \theta) Q(t_2, \theta) \cdots Q(t_n, \theta) dt_n \cdots dt_1, \quad (12)$$

where the right-hand member only depends on the intensities and where the series converges uniformly. The following theorem—akin to a multidimensional Gronwall-type inequality—is a natural result.

Theorem 5 (Dependence of the transition probabilities on the intensities). *Let $\|\cdot\|$ be a matrix norm and let $Q_1(t, \theta)$ and $Q_2(t, \theta)$ be two matrices of intensities norm bounded by $M > 0$ in $[0, T]$. Define*

$$\epsilon(Q_1, Q_2) := \sup_{t \in [0, T], \theta \in \Theta} \|Q_1(t, \theta) - Q_2(t, \theta)\|. \quad (13)$$

Then, we have that

$$\sup_{t \in [x, T], \theta \in \Theta} \|P_1(x, t, \theta) - P_2(x, t, \theta)\| \leq \epsilon(Q_1, Q_2) \frac{e^{M|T-x|} - 1}{M}, \quad (14)$$

where $P_1(x, u, \theta)$ and $P_2(x, u, \theta)$ are the solutions of the Kolmogorov equations—given in Formula (1)—with matrices of intensities $Q_1(t, \theta)$ and $Q_2(t, \theta)$, respectively.

Proof. The proof of a result of this type for an ordinary differential equation $y'(t) = f(t, y(t))$, satisfying the condition

$$|f(t, y_1(t)) - f(t, y_2(t))| \leq \lambda(t)|y_1(t) - y_2(t)|,$$

where λ is integrable and is immediate from the integral representation of the differential equation. So we will use the integral representation given by Formula (12). The following well-known result (see [38] p. 217 and, for a proof, [28] p. 348) will be instrumental.

Lemma 1. Let $q : \mathbb{R}_+ \mapsto \mathbb{R}$ a measurable function integrable over every bounded interval of \mathbb{R}_+ . Then, we have that

$$\int_s^t \int_{s_1}^t \cdots \int_{s_{n-1}}^t q(s_1)q(s_2) \dots q(s_n) ds_n \dots ds_2 ds_1 = \frac{\left(\int_s^t q(u) du\right)^n}{n!},$$

for all $0 \leq s \leq t, n \geq 1$.

Let us show, by induction, that if $\|Q_1(t, \theta)\| \leq M$ and $\|Q_2(t, \theta)\| \leq M$ for some $0 < M < +\infty$ then, using hypothesis in Formula (13), we have that

$$\begin{aligned} \left\| \prod_{k=1}^n Q_1(t_k, \theta) - \prod_{k=1}^n Q_2(t_k, \theta) \right\| &\leq \max_{k=1, \dots, n} \|Q_1(t_k, \theta) - Q_2(t_k, \theta)\| \cdot M^{n-1} \leq \\ &\leq \epsilon(Q_1, Q_2) M^{n-1} \end{aligned} \quad (15)$$

In fact, for the first order bound we have that

$$\begin{aligned} &Q_1(t_1, \theta)Q_1(t_2, \theta) - Q_2(t_1, \theta)Q_2(t_2, \theta) = \\ &= Q_1(t_1, \theta)Q_1(t_2, \theta) - Q_1(t_1, \theta)Q_2(t_2, \theta) + Q_1(t_1, \theta)Q_2(t_2, \theta) - Q_2(t_1, \theta)Q_2(t_2, \theta) = \\ &= Q_1(t_1, \theta)(Q_1(t_2, \theta) - Q_2(t_2, \theta)) + (Q_1(t_1, \theta) - Q_2(t_1, \theta))Q_2(t_2, \theta), \end{aligned}$$

and then it follows, using the matrix norm hypothesis, that

$$\begin{aligned} &\|Q_1(t_1, \theta)Q_1(t_2, \theta) - Q_2(t_1, \theta)Q_2(t_2, \theta)\| \leq \\ &\leq M\|Q_1(t_2, \theta) - Q_2(t_2, \theta)\| + M\|Q_1(t_1, \theta) - Q_2(t_1, \theta)\| \leq \\ &\leq M \cdot \max_{k=1,2} \|Q_1(t_k, \theta) - Q_2(t_k, \theta)\|. \end{aligned}$$

Consider now, for clarity, the next induction step, the second order bound.

$$\begin{aligned} &Q_1(t_1, \theta)[Q_1(t_2, \theta)Q_1(t_3, \theta)] - Q_2(t_1, \theta)[Q_2(t_2, \theta)Q_2(t_3, \theta)] = \\ &= Q_1(t_1, \theta)[Q_1(t_2, \theta)Q_1(t_3, \theta)] - Q_1(t_1, \theta)[Q_2(t_2, \theta)Q_2(t_3, \theta)] + \\ &+ Q_1(t_1, \theta)[Q_2(t_2, \theta)Q_2(t_3, \theta)] - Q_2(t_1, \theta)[Q_2(t_2, \theta)Q_2(t_3, \theta)] = \\ &= Q_1(t_1, \theta)([Q_1(t_2, \theta)Q_1(t_3, \theta)] - [Q_2(t_2, \theta)Q_2(t_3, \theta)]) + \\ &+ (Q_1(t_1, \theta) - Q_2(t_1, \theta))[Q_2(t_2, \theta)Q_2(t_3, \theta)]. \end{aligned}$$

Again it follows, from the matrix norm hypothesis and using the previous order one bound, that

$$\begin{aligned} & \|Q_1(t_1, \theta)Q_1(t_2, \theta)Q_1(t_3, \theta) - Q_2(t_1, \theta)Q_2(t_2, \theta)Q_2(t_3, \theta)\| \leq \\ & \leq M\|[Q_1(t_2, \theta)Q_1(t_3, \theta)] - [Q_2(t_2, \theta)Q_2(t_3, \theta)]\| + M^2\|Q_1(t_1, \theta) - Q_2(t_1, \theta)\| \leq \\ & \leq M^2 \cdot \max_{k=1,2,3} \|Q_1(t_k, \theta) - Q_2(t_k, \theta)\|. \end{aligned}$$

The induction proof is now cleared. Now, by using Formulas (12) and (15) and Lemma 1, we have the following bound for the norm of the difference of the two transition probability matrices.

$$\begin{aligned} & \|P_1(x, t, \theta) - P_2(x, t, \theta)\| \leq \\ & \leq \sum_{n=1}^{+\infty} \int_{[x,t]} \int_{[t_1,t]} \cdots \int_{[t_{n-1},t]} \left\| \prod_{k=1}^n Q_1(t_k, \theta) - \prod_{k=1}^n Q_2(t_k, \theta) \right\| dt_n \cdots dt_1 \leq \\ & \leq \frac{1}{M} \sum_{n=1}^{+\infty} \int_{[x,t]} \int_{[t_1,t]} \cdots \int_{[t_{n-1},t]} \epsilon(Q_1, Q_2) M^n dt_n \cdots dt_1 \leq \\ & \leq \frac{1}{M} \sum_{n=1}^{+\infty} \int_{[x,t]} \int_{[t_1,t]} \cdots \int_{[t_{n-1},t]} \epsilon(Q_1, Q_2)^{1/n} M \times \cdots \times \epsilon(Q_1, Q_2)^{1/n} M dt_n \cdots dt_1 = \\ & = \frac{1}{M} \epsilon(Q_1, Q_2) \sum_{n=1}^{+\infty} \frac{M^n |t-x|^n}{n!} = \epsilon(Q_1, Q_2) \frac{e^{M|T-x|} - 1}{M}, \end{aligned}$$

thus proving the result, in Formula (14). \square

A result like the one given by Theorem 5 is expected to simplify the estimation of the parameters $\theta \in \Theta$ that allow the fitting of a Markov chain model to real data coming, for example, from multi-state models for health insurance or long-term care, which was the case discussed in [1].

Remark 6 (Applying Theorem 5). *The applicability of Formula (14) requires that $M|T-x| \ll 1$; if not, the result is of no use. The usefulness of the result relies on the possibility of localising the computation of the solutions of Kolmogorov ODE. Once time units are chosen—let us say, one year—this is achieved by solving, successively, the Kolmogorov differential equations in time intervals $[x_k, x_{k+1}]$ of length $|x_{k+1} - x_k| \ll 1$. In doing so, the final values of the transition probability matrix in one interval must be the initial values of the transition probability matrix Kolmogorov ODE in the immediately following time interval. Supposing the adequate hypothesis for the existence and unicity of solutions of the Kolmogorov forward equations, we may then conclude that two matrix intensity matrices are close to one another in some small time interval, the correspondent probability transition matrices will be close to one another in the same small time interval.*

An illustrative example of the usefulness of this result is given in Figure 2 for which the intensity matrix considered is the following.

$$\begin{pmatrix} \mu_{11}(t) & 1.20135 \times 10^{-5} e^{0.117(t+50)} + \frac{1}{200} & 1.2 \cdot \mu_{12}(t) & 0.05 \cdot \mu_{34}(t) \\ 0.7 \cdot \mu_{12}(t) & \mu_{22}(t) & 5.49958 \times 10^{-6} e^{0.128(t+50)} + \frac{3}{500} & 0.5 \cdot \mu_{34}(t) \\ 0.36 \cdot \mu_{12}(t) & 0.3 \cdot \mu_{23}(t) & \mu_{33}(t) & 4.08902 \times 10^{-6} e^{0.139(t+50)} + \frac{7}{1000} \\ 0. & 0 & 0 & 1 \end{pmatrix} \quad (16)$$

That is all the intensities are of Gompertz–Makeham type; moreover, $\mu_{12}(t)$, $\mu_{23}(t)$ and $\mu_{34}(t)$ were first defined, and then, all the others were defined proportional to these three; the determinations of the parameters of these intensities from the data is the goal of an estimation calibration procedure. The coefficients of $\mu_{12}(t)$, $\mu_{23}(t)$ and $\mu_{34}(t)$ were chosen having in mind a unit time of one year and a LTC model starting at the age of 50 years and

going on until 100 years of age. The proportions between $\mu_{12}(t)$, $\mu_{23}(t)$ and $\mu_{34}(t)$ and all the others can be calibrated using a discrete-time transition matrix if available. The linear interpolations of the intensities were given at the six following points $t = 0, 15, 30, 40, 45, 50$. The differences of the linear interpolated intensities and the original Gompertz–Makeham intensities $\mu_{12}(t)$, $\mu_{23}(t)$ and $\mu_{34}(t)$ are shown in Figure 3.

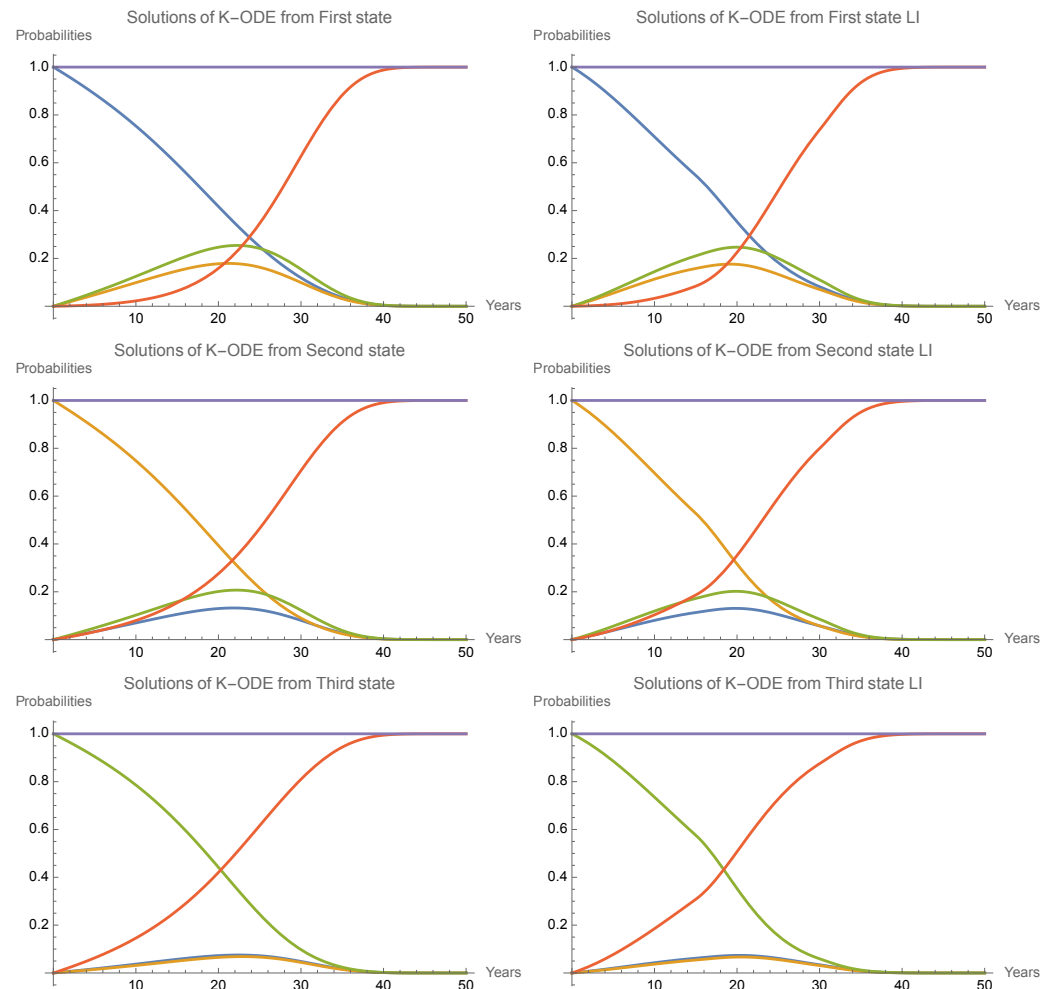


Figure 2. Comparing transition probabilities from Gompertz–Makeham intensities (left-hand side) and corresponding six-point linear interpolations (right-hand side). p_{ij} : $j = 1$ (blue); $j = 2$ (orange); $j = 3$ (green); $j = 4$ (red). $\sum_{j=1}^4 p_{ij}$ (purple).

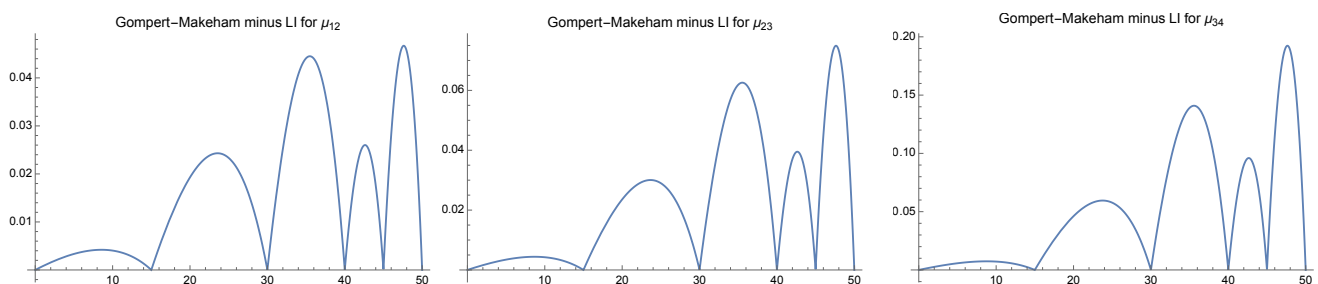


Figure 3. The symmetric of the differences between Gompertz–Makeham intensities and corresponding six-point linear interpolations for $\mu_{12}(t)$, $\mu_{23}(t)$ and $\mu_{34}(t)$.

The analysis of Figure 3 together with Figure 2 conveys an illustration of Theorem 5 and Remark 6 in the particular case of the approximation of Gompertz–Makeham intensities by linear interpolated ones.

4. Constructive Definition of CT-MC

The existence of a non-homogenous continuous-time Markov chain can also be guaranteed by a known constructive procedure that we now present, for completeness, and that is most useful for simulation and that we will use for defining the estimation calibration procedure proposed in this work. A reference for the following algorithmic definition of a Markov chain in continuous time is [37] (p. 266). For a proof of Theorem 6 below, see [38] (pp. 221–233). Let $\theta \in \Theta$ be a parameter.

Definition 3 (Constructive definition). *Given a transition intensity matrix,*

$$Q(t, \theta) = [\mu_{ij}^{\theta}(t)]_{i,j=1,\dots,d},$$

define

$$p^*(t, i, j) = \begin{cases} \frac{1 - \delta_i^j}{-\mu_{ii}^{\theta}(t)} \mu_{ij}^{\theta}(t) & \mu_{ii}^{\theta}(t) \neq 0 \\ \delta_i^j & \mu_{ii}^{\theta}(t) = 0, \end{cases} \quad (17)$$

with δ_i^j Kronecker's delta. Let $X_0 = i$, according to some initial distribution on $\{1, 2, \dots, d\}$.

1. The jump sequence $(\tau_n)_{n \geq 0}$ of stopping times is defined by induction as follows; $\tau_0 \equiv 0$.
2. τ_1 , the sojourn time in state i which is also the time of first jump, has an exponential distribution function given by

$$F_{\tau_1}(t) = \mathbb{P}[\tau_1 \leq t] = 1 - \exp\left(-\int_0^t \mu_{ii}^{\theta}(u) du\right). \quad (18)$$

We note that this distribution of the stopping time is mandatory as a consequence of a general result on the distribution of sojourn times of a continuous-time Markov chain (see Theorem 2.3.15 in [38] p. 221).

3. Given that the process is in state i , it may jump to state j at time $\tau_1 = s_1$ with probability $p^*(t, i, j)$ defined in Formula (3), that is

$$\mathbb{P}[X_{s_1} = j | \tau_1 = s_1, X_0 = i] = p^*(s_1, i, j), \quad (19)$$

and so $X_t = i$ for $0 \equiv \tau_0 \leq t < \tau_1$.

4. Given that $\tau_1 = s_1$ and $X_{s_1} = j$, τ_2 time of the second jump with an exponential distribution function,

$$F_{\tau_2|\tau_1=s_1}(t) = \mathbb{P}[\tau_2 \leq t | \tau_1 = s_1] = 1 - \exp\left(-\int_0^t \mu_{jj}^{\theta}(u + s_1) du\right)$$

and

$$\mathbb{P}[X_{s_2} = k | \tau_1 = s_1, X_0 = i, \tau_2 = s_2, X_{s_1} = j] = p^*(s_1 + s_2, j, k),$$

and so $X_t = j$ for $\tau_1 \leq t < \tau_2$.

The following theorem ensures that the preceding construction yields the desired result.

Theorem 6 (The continuous-time Markov chain). *Let the intensity matrix be norm bounded by a Lebesgue integrable function in $[0, T]$. Then, given the times $(\tau_0)_{n \geq 1}$, we have that with the sequence $(Y_n)_{n \geq 1}$ defined by $Y_n = X_{\tau_n}$, the process defined by*

$$X_t = \sum_{n=0}^{+\infty} Y_n \mathbb{1}_{[\tau_n, \tau_{n+1}[}(t) = \sum_{n=0}^{+\infty} X_{\tau_n} \mathbb{1}_{[\tau_n, \tau_{n+1}[}(t) \quad (20)$$

which is a continuous-time Markov chain with transition probabilities P and transition intensities Q .

Proof. This theorem is stated and proved, in the general case of continuous-time Markov processes in [38] (p. 229). \square

5. Estimation–Calibration Procedure

In the following, we consider the set of procedures that allow us to obtain transition intensities from simulated data and then, by integration of the forward Kolmogorov ODE, the transition probabilities; these so-called estimated transition probabilities will be compared to the original transition probabilities that were used to generate the simulated data. The procedures comprehend both non-parametric statistical estimation by kernel methods and fitting piecewise linear functions to data with additional constraints, a procedure more akin to calibration.

We will consider an ideal sample of complete data represented in Figure 4; each line represents a *trajectory* and we have on the left-hand side the initial state, following it, we have the time spent in that state then the new state and the time spent in that state and so on and so forth...

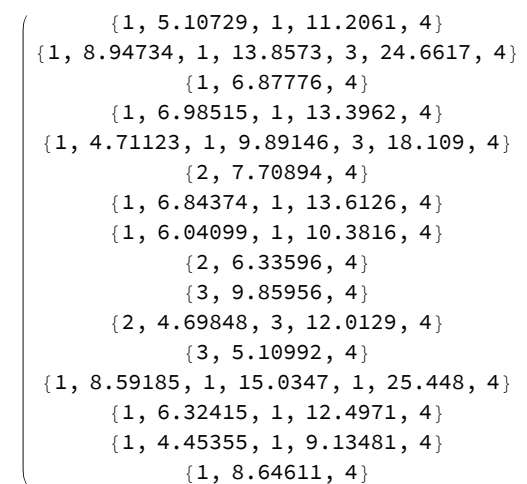


Figure 4. A set of simulated trajectories of a 4-state continuous-time Markov chain.

Using the procedure detailed in Section 4, these data were generated with a full intensity matrix, that is a matrix of the form,

$$\begin{pmatrix} -(\mu_{12} + \mu_{13} + \mu_{14}) & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & -(\mu_{21} + \mu_{23} + \mu_{24}) & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & -(\mu_{31} + \mu_{32} + \mu_{34}) & \mu_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (21)$$

with the intensities given by

$$\begin{cases} \mu_{12} = 3.47 \cdot 10^{-6} e^{0.138(t+65)} + \frac{1}{2500} = \mu_{21} & \mu_{13} = 0.5 \cdot \mu_{12} & \mu_{23} = 1.5 \cdot \mu_{12} \\ \mu_{14} = 0.0000758 e^{0.087(t+65)} + \frac{1}{2000} & \mu_{24} = 1.4 \cdot \mu_{14} & \mu_{34} = 1.8 \cdot \mu_{14} \\ \mu_{21} = \mu_{12} & \mu_{31} = 0.1 \cdot \mu_{21} & \mu_{32} = 0.4 \cdot \mu_{21} \end{cases} \quad (22)$$

This set of intensities used to generate the full data sample was supposed to determine a model for LTC with four states; the relations between the intensities reflect the qualitative relations describing the force of transitions that we suppose are held in this particular model.

For estimation–calibration purposes—following the results on the distance of transition probabilities—we will suppose that it has given the most tractable functional form of

the intensities depending on some parameters to be estimated. For instance, a continuous piecewise linear functional form for which we can have, for $i \neq j$,

$$\mu_{ij}^{\theta}(t) = \sum_{k=1}^r \left(\theta_{ij}^{k,1} + \theta_{ij}^{k,2} t \right) \mathbb{1}_{[t_k, t_{k+1}[}(t), \quad (23)$$

with $0 = t_0 < t_1 < \dots < t_{r+1} = T$, and possibly some conditions on the parameters $\theta_{ij}^{k,1}$ and $\theta_{ij}^{k,2}$ if the intensity $\mu_{ij}^{\theta}(t)$ is supposed to be piecewise linear and continuous.

We stress again that the values in Figure 4 were simulated. For LTC, for example, real data will be given, possibly, with time stamps of a day, at best; being so, the order of approximation will have to be chosen looking for precision with a balance between a sufficiently narrow interval around a given time and having enough observations to perform the estimation.

Let us detail a methodology to identify the parameter $\theta \in \Theta$ inspired by the constructive definition of the Markov chain in Remark 3. The general idea of the methodology is as follows.

- (i) Given a state i , we have to find a fitting for the distribution of random times of $i \rightarrow i$ jumps. According to Formula (18), these times have an exponential distribution with density $\mu_{ii}^{\theta}(t)$.
- (ii) For every other state j , by using $\mathbb{P}[X_{s_1} = j | \tau_1 = s_1, X_0 = i]$, possibly with an approximation, by Formula (19), we can obtain an approximation of $p^*(s_1, i, j)$.
- (iii) By using Formula (17) and the approximation obtained for $p^*(s_1, i, j)$, we can obtain an approximation for $\mu_{ij}^{\theta}(t)$.
- (iv) Finally, we will fit a linear continuous piecewise intensity to $\mu_{ij}^{\theta}(t)$.

Let us detail the procedures for applying the methodology just described.

1. Recall that an observed trajectory has the following structure: (first state, time spent in state, second state, time spent in state, third state ...). The maximum length of a trajectory in our sample is 11. Select all the trajectories of length greater than 3 that start at state $i = 1$. If the next state is also $i = 1$, the time spent in state—in this case, in state $i = 1$ —is the first part of the sample for obtaining $\mu_{11}^{\theta}(t)$. Select all the trajectories of length greater than 5 for which the second state is $i = 1$; this set of trajectories already contains the previous considered set of trajectories and so, if the third state is also $i = 1$, the sum of the time spent in the first state and the time spent in the second state will be the second part of the sample for obtaining $\mu_{11}^{\theta}(t)$. Repeat, successively, the procedure for all trajectories of length greater than 7, then of length greater than 9 and finally of length greater than 11 to obtain the full sample for the intensity $\mu_{11}^{\theta}(t)$.
2. Fit a smooth kernel distribution to the sample obtaining the intensity $\mu_{11}^{\theta}(t)$.
3. Repeat the procedure used for obtaining the sample for $\mu_{11}^{\theta}(t)$, but this time, select the transitions $1 \rightarrow 2$, that is, the transitions from state $i = 1$ to state $i = 2$. Fit a smooth kernel distribution to these data.
4. Now, we look for an estimate of $p^*(t, i, j)$ given by Formulas (17) and (19). For that, we will consider rounding the sojourn times—say to the unity, in order to have enough observations—and then group all observations of jumps from the first state according to this rounding. Consider then the observations towards state $i = 2$. We will then have that

$$\begin{aligned} p^*(s_1, 1, 2) &= \mathbb{P}[X_{s_1} = 2 | \tau_1 = s_1, X_0 = 1] \approx \frac{\mathbb{P}[X_{s_1} = 2, s_1 - 0.5 \leq \tau_1 < s_1 + 0.5]}{\mathbb{P}[s_1 - 0.5 \leq \tau_1 < s_1 + 0.5]} = \\ &= \frac{\mathbb{P}[s_1 - 0.5 \leq \tau_1 < s_1 + 0.5 | X_{s_1} = 2] \cdot \mathbb{P}[X_{s_1} = 2]}{\mathbb{P}[s_1 - 0.5 \leq \tau_1 < s_1 + 0.5]} \end{aligned} \quad (24)$$

and most of the left-hand side of the formula will be estimable with the observations by using the smooth kernel distributions.

5. Resorting to Formula (17), we can compute values for $\mu_{12}^{\theta}(t)$ and fit a piecewise linear density. That is, using again Formula (17), since $\mu_{11}^{\theta}(s_1) \neq 0$, we have that for an arbitrary time $t = s_1$,

$$\mu_{12}^{\theta}(s_1) \approx \frac{-\mu_{11}^{\theta}(s_1)}{1 - \delta_1^2} p^*(s_1, 1, 2) = -\mu_{11}^{\theta}(s_1) \cdot p^*(s_1, 1, 2).$$

6. We consider a set of values of $\mu_{12}^{\theta}(s_1), \mu_{12}^{\theta}(s_2), \dots, \mu_{12}^{\theta}(s_k)$ and then fit the multidimensional parameter $\theta \in \Theta$ to these values (see Formula (23) for the case of a continuous piecewise linear intensity functional form).
7. These procedures are to be repeated in order to obtain the parameters μ_{2j}^{θ} for $j \neq 2$ and μ_{3j}^{θ} for $j \neq 3$.
8. The intensities μ_{jj}^{θ} for $j = 1, 2, 3$ are obtained in the usual form and are forcible continuous piecewise linear since they are the sum of continuous piecewise linear functions.

6. Results of the Estimation–Calibration Procedure

We present the results from the estimation procedure developed in accordance with the methodology proposed in Section 5.

The estimated matrix structure is a full matrix such as the one given in Formula (21) but with piecewise linear intensities which are not very elucidative in themselves. It is preferable to look at a graphical representation of these intensities. In Figure 5, we have the estimated intensities (in blue colour) and the fitted piecewise linear intensities (in red colour). We can observe that despite observable differences the fitting is reasonably good with the exception of the intensity μ_{21} . This may be due to the fact that we only had seven observations in the sample, and they are consistent with LTC data. We also observe that the fitting with an exponential-type density will give a non satisfactory result. In order to evaluate the quality of this fitting, we present in Table 1 the maximum distance between the computed approximate intensities and the fitted piecewise linear intensities.

Table 1. (1)—Maximum distance between the values of the computed approximate intensities; and (2)—the fitted piecewise linear intensities and the maximum distance normalised by the maximum value of the estimated intensities.

	μ_{12}	μ_{13}	μ_{14}
(1)	0.025381	0.159100	1.165812
(2)	0.000667	0.004186	0.030679
	μ_{21}	μ_{23}	μ_{24}
(1)	0.003661	0.020507	0.397256
(2)	0.000963	0.005396	0.104541
	μ_{31}	μ_{32}	μ_{34}
(1)	7.398272	0.000024	0.086817
(2)	3.52299	0.000011	0.04134

We observe that the error for μ_{31} is quite large compared to the other errors; this could be due to the fact that the estimation was performed with only 20 observations, again consistent with LTC data.

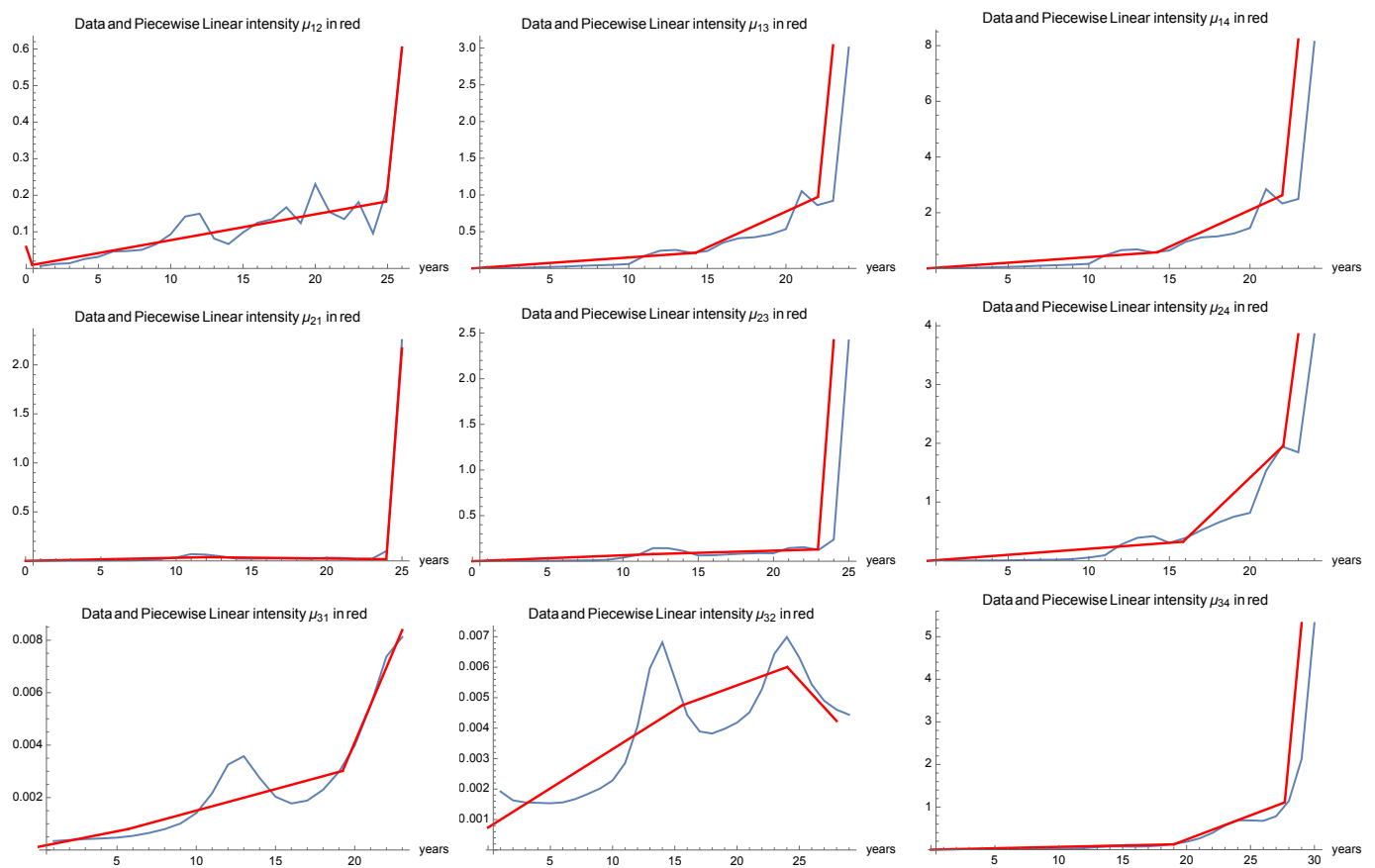


Figure 5. Data recovered from the simulated results (blue) and the piecewise linear intensities fitted (red).

In Figure 6, we can compare qualitatively the original transition probabilities with the ones obtained as the solution of the forward Kolmogorov ODE with the estimated piecewise linear intensities. A first qualitative observation is that the general behaviour of the estimated and fitted intensities is similar. In order to compare quantitatively the original transition probabilities with the ones obtained as the solution of the Kolmogorov ODE with the estimated piecewise linear intensities, we present, in Figure 7, the difference between the original transition probabilities and the estimated transition probabilities for each of the three transient states. It is clear that there are substantial differences. To justify these differences we have at least two cumulative sources of error. The first error source is induced by the fact that there was an estimation–calibration procedure applied to 5000 trajectories generated from the original transition probabilities; the second error source comes from the fact that while the original transition probabilities are produced from intensities of the Gompertz–Makeham type, the estimated transition probabilities are produced by continuous piecewise linear intensities fitted to the simulated data.

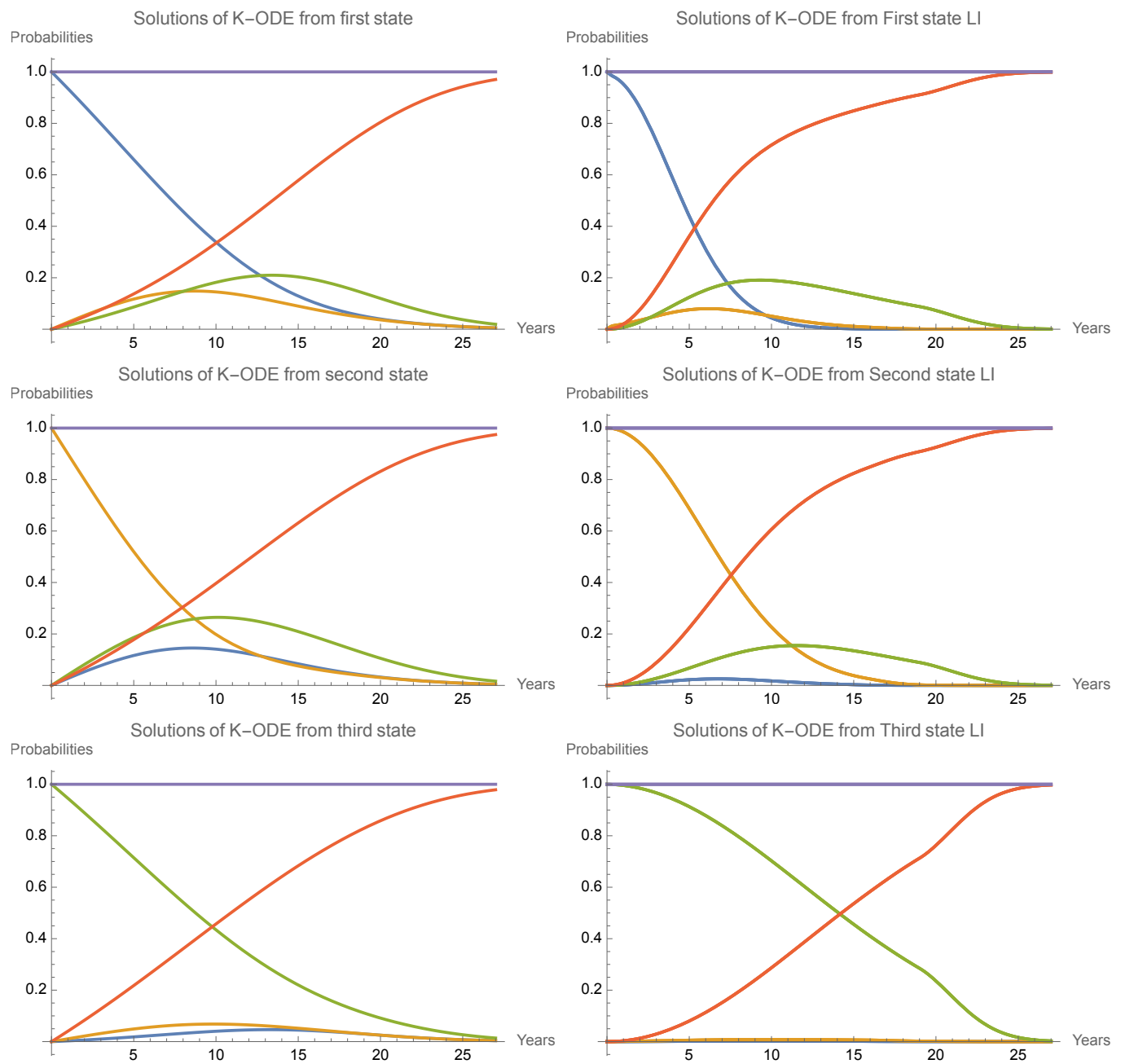


Figure 6. Comparing original transition probabilities used for simulation (left) with the estimated piecewise linear transition probabilities (right). p_{ij} : $j = 1$ (blue); $j = 2$ (orange); $j = 3$ (green); $j = 4$ (red). $\sum_{j=1}^4 p_{ij}$ (purple).

In order to detail the average error per year coming from the estimation procedure we can compute

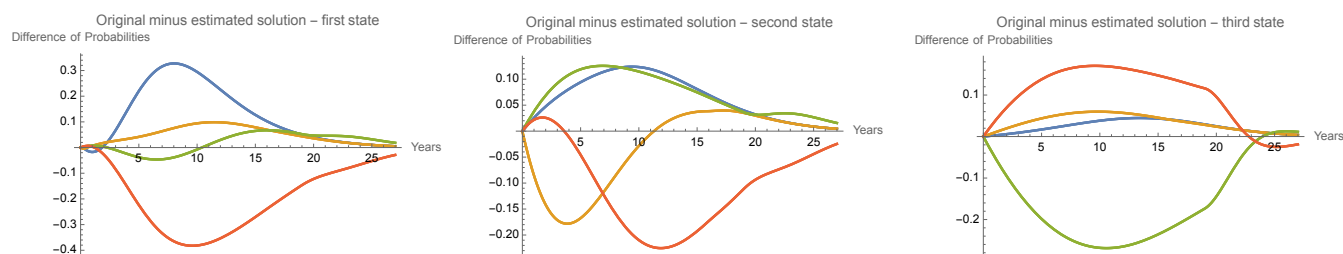
$$\Delta_{ij} := \frac{1}{27} \int_0^{27} |p_{ij}(t) - \tilde{p}_{ij}(t)| dt \quad (25)$$

with $p_{ij}(t)$ the original transition probabilities and $\tilde{p}_{ij}(t)$ the estimated transition probabilities. The results are presented in Table 2.

Table 2. Average error per year between the estimated and the original probability transitions.

Δ_{11}	Δ_{12}	Δ_{13}	Δ_{14}
0.004647	0.001806	0.000896	0.007118
Δ_{21}	Δ_{22}	Δ_{23}	Δ_{24}
0.002394	0.002140	0.002624	0.004081
Δ_{31}	Δ_{32}	Δ_{33}	Δ_{34}
0.000896	0.001258	0.005907	0.003898

The main conclusion is that the order of magnitude of the average error per unit time—chosen as a year since it is the time duration of an interval where transition probabilities in health insurance and long-term-care models can have a significant impact—is always less than 1%. Of course, we have to be careful of the extremes of these errors that are visible in Figure 7.

**Figure 7.** The difference between the original transition probabilities and the estimated ones. $p_{ij} - \tilde{p}_{ij}$: $j = 1$ (blue); $j = 2$ (orange); $j = 3$ (green); $j = 4$ (red).

All computations and graphic representations were created with a Mac mini (M1 2020) equipped with macOS Monterey 12.5.1 with Wolfram Mathematica 12, version 12.3.1.0. The estimation–calibration procedures use either native functions or small routines that require reasonable execution times of the order of a second.

7. Discussion

The methodology proposed in this work gave us the continuous piecewise linear intensities depicted in Figure 5. The use of the continuous piecewise linear functional form was intentional although not necessary; a better fit to the data, most possibly with a larger number of parameters, could be computed and possibly could provide a better final result, qualitatively, in Figure 6 and quantitatively with metrics given by both the average error per year, as in Table 2 and as in Figure 7, showing the analysis of local discrepancies between the original transition probabilities and the transition probabilities resulting from the estimation methodology. The intention of using the continuous piecewise linear functional form for the intensities was to illustrate the possibility of an estimation–calibration procedure relying on a small number of parameters. Whenever faced with the estimation–calibration of intensities for real-data modelling, a situation where there are no known determined intensities generating the data, the choice of the functional form is secondary with respect to the quality of the model fitting.

8. Conclusions

In this work, we proposed a methodology for estimation–calibration of continuous-time non-homogeneous Markov chains with finite state space. We presented an application of the methodology to a Monte Carlo simulated set of trajectories generated from intensities of Gompertz–Makeham type, and we obtained, by the methodology, estimated continuous piecewise linear intensities. We compared the correspondent transition probabilities—obtained by solving the forward Kolmogorov ODE for the original

Gompertz–Makeham intensities and for the continuous piecewise linear intensities—obtaining an average error per year of always less than 1%.

In order to justify the methodology, we presented a result on regime-switching Markov chains, proving the existence of a Markov chain process, in a given time interval, obtained by glueing together different intensities matrices defined in the different intervals of a partition of the time interval of the Markov chain process; this result shows that it is possible to consider intensities of different functional forms for different sub intervals of the time interval where the whole Markov chain process is defined.

We also presented a result that shows that it is possible to bound the distance between the matrix of transition probabilities corresponding to different matrix intensities by the distance between these matrix intensities. This result shows that, with respect to the transition probabilities, we can consider changes in functional form of the intensities—in an intensity matrix—as long as the distance between the original intensity matrix and the altered one is small enough.

In future work, we intend to improve the methodology in order to control the quality of the process by adequate tests and to improve the algorithm used.

Author Contributions: Conceptualization, M.L.E.; methodology, M.L.E., N.P.K. and G.R.G.; software, M.L.E., N.P.K. and G.R.G.; validation, M.L.E., N.P.K. and G.R.G.; formal analysis, M.L.E., N.P.K. and G.R.G.; investigation, M.L.E., N.P.K. and G.R.G.; data curation, M.L.E., N.P.K. and G.R.G.; writing—original draft preparation, M.L.E.; writing—review and editing, M.L.E., N.P.K. and G.R.G.; supervision, M.L.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by national funds through the FCT—Fundação para a Ciência e Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020—Center for Mathematics and Applications.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors are indebted to Albert N. Shiryaev for drawing attention to very important work on the subject and to the referees for the improvement suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MCCT	Markov chain in continuous time
ODE	ordinary differential equations

References

- Esquivel, M.; Guerreiro, G.; Oliveira, M.; Real, P. Calibration of Transition Intensities for a Multistate Model: Application to Long-Term Care. *Risks* **2021**, *9*, 37. [\[CrossRef\]](#)
- Vassiliou, P.C.G. *Non-Homogeneous Markov Chains and Systems—Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2023; pp. xxi+450. [\[CrossRef\]](#)
- Pitacco, E. *Health Insurance*; European Actuarial Academy (EAA) Series; Basic Actuarial Models; Springer: Cham, Switzerland, 2014; pp. xii+162. [\[CrossRef\]](#)
- Haberman, S.; Pitacco, E. *Actuarial Models for Disability Insurance*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1999; pp. xx+280.
- Billingsley, P. *Statistical inference for Markov Processes*; Volume II, *Statistical Research Monographs*; University of Chicago Press: Chicago, IL, USA, 1961; pp. vii+75.
- Billingsley, P. Statistical methods in Markov chains. *Ann. Math. Statist.* **1961**, *32*, 12–40. [\[CrossRef\]](#)
- Waters, H.R. An Approach to the Study of Multiple State Models. *J. Inst. Actuar. (1886–1994)* **1984**, *111*, 363–374. [\[CrossRef\]](#)
- Wolthuis, H. *Life Insurance Mathematics (The Markovian Model)*; CAIRE: Brussels, Belgium, 1994.
- Forfar, D.O.; McCutcheon, J.J.; Wilkie, A.D. On Graduation by Mathematical Formula. *J. Inst. Actuar. (1886–1994)* **1988**, *115*, 1–149. [\[CrossRef\]](#)
- Dickson, D.C.M.; Hardy, M.R.; Waters, H.R. *Actuarial Mathematics for Life Contingent Risks*, 3rd ed.; International Series on Actuarial Science; Cambridge University Press: Cambridge, UK, 2020; pp. xxiv+759. [\[CrossRef\]](#)
- Christiansen, M.C. Multistate models in health insurance. *ASTA Adv. Stat. Anal.* **2012**, *96*, 155–186. [\[CrossRef\]](#)

12. Wang, B.; Zhu, Q. Stability Analysis of Discrete-Time Semi-Markov Jump Linear Systems. *IEEE Trans. Autom. Control.* **2020**, *65*, 5415–5421. [\[CrossRef\]](#)
13. Wang, B.; Zhu, Q.; Li, S. Stability Analysis of Discrete-Time Semi-Markov Jump Linear Systems with Time Delay. *IEEE Trans. Autom. Control.* **2023**, *68*, 6758–6765. [\[CrossRef\]](#)
14. Meira-Machado, L.; de Uña Álvarez, J.; Cadarso-Suárez, C.; Andersen, P.K. Multi-state models for the analysis of time-to-event data. *Stat. Methods Med. Res.* **2009**, *18*, 195–222. [\[CrossRef\]](#)
15. Keiding, N.; Andersen, P.K. Nonparametric estimation of transition intensities and transition probabilities: A case study of a two-state Markov process. *J. R. Statist. Soc. Ser. C* **1989**, *38*, 319–329. [\[CrossRef\]](#)
16. Andersen, P.K.; Hansen, L.S.; Keiding, N. Non- and semi-parametric estimation of transition probabilities from censored observation of a nonhomogeneous Markov process. *Scand. J. Statist.* **1991**, *18*, 153–167.
17. Renshaw, A.E.; Haberman, S. On the graduations associated with a multiple state model for permanent health insurance. *Insur. Math. Econ.* **1995**, *17*, 1–17. [\[CrossRef\]](#)
18. Fong, J.H.; Shao, A.W.; Sherris, M. Multistate actuarial models of functional disability. *N. Am. Actuar. J.* **2015**, *19*, 41–59. [\[CrossRef\]](#)
19. Wan, L.; Lou, W.; Abner, E.; Kryscio, R.J. A comparison of time-homogeneous Markov chain and Markov process multi-state models. *Commun. Stat. Case Stud. Data Anal. Appl.* **2016**, *2*, 92–100. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Naka, P.; Boado-Penas, M.d.C.; Lanot, G. A multiple state model for the working-age disabled population using cross-sectional data. *Scand. Actuar. J.* **2020**, *2020*, 700–717. [\[CrossRef\]](#)
21. Qian, C.; Srivastava, D.K.; Pan, J.; Hudson, M.M.; Rai, S.N. Estimating transition intensity rate on interval-censored data using semi-parametric with EM algorithm approach. In *Communications in Statistics—Theory and Methods*; Taylor and Francis: Abingdon, UK, 2023; pp. 1–17. [\[CrossRef\]](#)
22. de Mol van Otterloo, S.; Alonso-García, J. A multi-state model for sick leave and its impact on partial early retirement incentives: The case of the Netherlands. *Scand. Actuar. J.* **2023**, *2023*, 244–268. [\[CrossRef\]](#)
23. Esquivel, M.; Patrício, P.; Guerreiro, G. From ODE to Open Markov Chains, via SDE: An application to models for infections in individuals and populations. *Comput. Math. Biophys.* **2020**, *8*, 180–197. [\[CrossRef\]](#)
24. Esquivel, M.; Krasii, N.; Guerreiro, G.; Patrício, P. The Multi-Compartment SI(RD) Model with Regime Switching: An Application to COVID-19 Pandemic. *Symmetry* **2021**, *13*, 2427. [\[CrossRef\]](#)
25. Feinberg, E.A.; Mandava, M.; Shiryayev, A.N. On solutions of Kolmogorov’s equations for nonhomogeneous jump Markov processes. *J. Math. Anal. Appl.* **2014**, *411*, 261–270. [\[CrossRef\]](#)
26. Feinberg, E.; Mandava, M.; Shiryayev, A.N. Kolmogorov’s equations for jump Markov processes with unbounded jump rates. *Ann. Oper. Res.* **2022**, *317*, 587–604. [\[CrossRef\]](#)
27. Feinberg, E.A.; Shiryayev, A.N. Kolmogorov’s equations for jump Markov processes and their applications to control problems. *Theory Probab. Appl.* **2022**, *66*, 582–600. *Transl. Teor. Veroyatn. Primen.* **2021**, *66*, 734–759. [\[CrossRef\]](#)
28. Rolski, T.; Schmidli, H.; Schmidt, V.; Teugels, J. *Stochastic Processes for Insurance and Finance*; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd.: Chichester, UK, 1999; pp. xviii+654. [\[CrossRef\]](#)
29. Coddington, E.A.; Levinson, N. *Theory of Ordinary Differential Equations*; McGraw-Hill Book Co., Inc.: New York, NY, USA, 1955; pp. xii+429.
30. Hale, J.K. *Ordinary Differential Equations*, 2nd ed.; Robert E. Krieger Publishing Co., Inc.: Huntington, NY, USA, 1980; pp. xvi+361.
31. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2013; pp. xviii+643.
32. Stokes, A. The application of a fixed-point theorem to a variety of nonlinear stability problems. *Proc. Nat. Acad. Sci. USA* **1959**, *45*, 231–235. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Kurzweil, J. Ordinary differential equations. In *Studies in Applied Mechanics*; Elsevier Scientific Publishing Co.: Amsterdam, The Netherlands, 1986; Volume 13, 440p.
34. Teschl, G. *Ordinary Differential Equations and Dynamical Systems; Graduate Studies in Mathematics*; American Mathematical Society: Providence, RI, USA, 2012; Volume 140, pp. xii+356. [\[CrossRef\]](#)
35. Nevanlinna, F.; Nevanlinna, R. *Absolute Analysis*; Springer: New York, NY, USA, 1973; Volume 102, pp. vi+270.
36. Butcher, J.C. *Numerical Methods for Ordinary Differential Equations*, 3rd ed.; John Wiley & Sons, Ltd.: Chichester, UK, 2016; pp. xxiii+513. [\[CrossRef\]](#)
37. Iosifescu, M. *Finite Markov Processes and Their Applications*; (Wiley Series in Probability and Mathematical Statistics); John Wiley & Sons, Ltd.: Chichester, UK; Editura Tehnică: Bucharest, Romania, 1980; p. 295.
38. Iosifescu, M.; Tăutu, P. *Stochastic Processes and Applications in Biology and Medicine. I: Theory*; Editura Academiei RSR: Bucharest, Romania; Springer: Berlin/Heidelberg, Germany, 1973; Volume 3, p. 331.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.