mathematics

Article

Calendar Effect and In-Sample Forecasting Applied to Mesothelioma Mortality Data

Alex Isakson [1,*], Simone Krummaker [1], María Dolores Martínez-Miranda [2] and Ben Rickayzen [1]

1   Bayes Business School (Formerly Cass), City, University of London, London EC1Y 8TZ, UK; simone.krummaker@city.ac.uk (S.K.); B.D.Rickayzen@city.ac.uk (B.R.)
2   Department of Statistics and Operations Research, University of Granada, 18071 Granada, Spain; mmiranda@ugr.es
*   Correspondence: Alex.Isakson@cass.city.ac.uk

Abstract: In this paper, we apply and further illustrate a recently developed extended continuous chain ladder model to forecast mesothelioma deaths. Making such a forecast has always been a challenge for insurance companies as exposure is difficult or impossible to measure, and the latency of the disease usually lasts several decades. While we compare three approaches to this problem, we show that the extended continuous chain ladder model is a promising benchmark candidate for asbestosis mortality forecasting due to its flexible and simple forecasting strategy. Furthermore, we demonstrate how the model can be used to provide an update for the forecast of the number of deaths due to mesothelioma in Great Britain using in recent Health and Safety Executive (HSE) data.

Keywords: continuous chain ladder; age-period-cohort model; backfitting; density estimation; kernel smoothing

JEL Classification: C14; C53

1. Introduction

1.1. Motivation

Since the 1960s, mesothelioma or asbestos-related cancer has gained worldwide interest as a result of its increasing incidence, related medico-legal issues and poor prognosis. Mesothelioma is mainly caused by occupational exposure to asbestos fibres in sectors such as mining, road, railway and general construction as well as shipyards, etc., which have a mainly male workforce [1]. A brief, or even indirect, exposure to a small dose of asbestos fibres might be enough to trigger the disease much later in life [2]. For example, the latency period of mesothelioma is between 20 and 70 years with an average of around 40 years. Once the symptoms appear, it is rapidly fatal, with the majority of deaths occurring amongst those over 60 years of age [1]. In Great Britain, mesothelioma mortality has been steadily increasing in recent years, with 2101 deaths recorded in 2016 and a trend for the average age of deaths to slowly increase over time. Notifications of mesothelioma claims have exhibited a stable but increasing trend so far [3].

Asbestos-related claims have a lasting impact on the global insurance industry. The industry is paying, on average, $1.9 billion for mesothelioma claims annually (2013–2017, [4]) under policies that covered, for example, employer or product liability at the time of exposure. This has resulted in multiple insurer insolvencies since the 1940s. Still, today, it is difficult to project the industry's ultimate loss exposure due to advances in treatment, increasing life expectancies, changes in litigation and the number of new claimants emerging. A core uncertainty for the insurance industry is, therefore, whether the amounts of technical reserves set aside to cover future claims are sufficient. The UK insurance market estimates are based on population mesothelioma deaths projected by the Health and Safety

Executive (HSE), which is the UK independent regulator with respect to health and safety in the workplace.

With respect to mesothelioma mortality forecasting, there are two key questions for insurers to answer: (i) when the numbers of deaths are going to peak (and establishing the peak value); and (ii) how the deaths will develop after the peak, that is, the shape of the forecasts. As insurers have to set aside reserves for future claims payments, the ultimate claim amounts and shape of how the number of deaths due to mesothelioma will reduce over time are of critical importance. If the shape is incorrectly forecasted, reserves might be overestimated or underestimated.

1.2. Literature Review

For decades, mortality forecasting has been an important tool for decision making in many fields such as actuarial science, economics, epidemiology and demography, to name only a few. With such great interest in the topic, numerous mortality forecasting approaches have been developed. The literature on mortality forecasting approaches is large, and we do not aim to provide a full review here (see for example [5] for a recent description). The first step in understanding and forecasting mortality patterns is to construct a model describing observed death counts or mortality rates, across age groups or within cohorts. The Lee–Carter model [6] is the current benchmark in mortality studies used by government agencies and pension funds. The model assumes that the dynamics of the logarithms of the central death rates are driven by an age specific constant plus the speed of change at each age multiplied by an overall time trend of mortality rates. The model has many extensions in the literature, which provide improved estimation procedures [7], a relaxation of assumptions [8] and adjustments to the model [9–11], among others.

Standard mortality approaches rest on dose–response analyses where both death counts and exposure are available. However, this is not the case in mesothelioma mortality forecasting, where the number of people who have been exposed to asbestos is unknown. There are two possible approaches to this problem. The first one is to construct a synthetic measure for exposure and to use a dose–response model. The second approach is to model only the observed number of deaths. The first approach has been used in the UK by the Health and Safety Executive (HSE), using a birth-cohort model [12–14], which assumes that the risk of mesothelioma depends on age and years of exposure, and that an individual's asbestos exposure is related to the year of exposure. However, a key problem with this approach is that the affected individuals could have been silently exposed to asbestos over prolonged periods of time so that there is no reliable measure for the exposure to asbestos. Therefore, the estimates include a high degree of uncertainty, and regular adjustments are necessary.

A discussion on modelling mortality with synthetic exposure can be found in [15]. Other approaches involving synthetic exposure include [16], who describes an application of the Lee–Cater model to forecast mesothelioma mortality in Argentina, while [17] considered Generalised Interactive Linear Models for Italian data, and [18] used Generalised Additive Models for Brazil. A different and simpler approach to model the observed number of deaths has been proposed by [15,19]. It does not rely on exposure measures and, therefore, avoids the difficulties associated with extrapolating them into the future. The approach is inspired by the so-called "chain ladder" method introduced by [20] in actuarial science. This is a technique used to calculate the liabilities in the form of outstanding claims faced by an insurance company. For an overview of the classical chain ladder method, see [21]. While the method was introduced as a deterministic algorithm, it is has been shown that it consists of an age-period-cohort (APC) model estimated using Poisson regression [22].

APC models have been studied for a long time. Refer to [23] for a general review of age-period-cohort models, and [24] for a recent review and comparison for cancer studies. They are primarily descriptive tools for data in Lexis diagrams when there are non-trivial age, period and cohort effects. A key issue in APC models is the overparameterisation

induced by the relation $age = period - cohort$. The consequence of this relationship is that, without further assumptions, APC models are not uniquely identifiable, i.e., the model has infinitely many fits with infinitely many interpretations [25]. The modelling challenges that come with this problem have been well formulated by [26,27] in discrete time and by [28] in continuous time. When facing overparametrisation, there are two choices, either to work with the over-parametrised model, using for instance an ad hoc identification of the parameters in order to specify them uniquely, or to use a unique and well-defined parametrisation based on a maximal invariant (in generalized linear models, this is the canonical parameter). Following the second approach, [22] proposed second differences in order to uniquely parametrise APC models. While the canonical parametrisation has important theoretical advantages, it might feel less intuitive to many researchers and applied analysts. Therefore, we can find many examples in the literature where non-unique parametrisations are used. Some additional issues arise when looking at models with non-linear parametrisations, such as the Lee–Carter model—see [29,30] for a discussion on the identification problem in this model. Other common approaches in mortality analysis include Bayesian methods or random effects methods (e.g., [31,32]). Unfortunately the identification problem remains in these cases—see [30] for a discussion on additional issues with these methods.

Usually, the objective of a mortality study is to forecast the future mortality, and extrapolation is often used. A major challenge when using extrapolative approaches is to identify the underlying long-term mortality trend that can be extrapolated. Unfortunately, the data might not contain enough information, and a careful analysis of past mortality and its determinants is often required. Within the APC models, the period and cohort parameters are typically treated as independent time series processes, which are used to project the parameters stochastically into the future. Reference [33] combines the canonical parametrisation of [22] with standard methods for the forecasting of non-stationary time series. Reference [34] considers Box–Jenkins procedures to determine the time series processes generating the parameters. Recently, [35,36] defined the term "in-sample forecasting" to mean forecasting a structured function in regions where the function is not observed, but where it is determined by its values in the observed region. This has several advantages over methods based on time series analysis, as [37] discussed. Standard APC models can be understood as discrete density models with a simple multiplicative structure (or log lineal). There are many examples in actuarial science and epidemiology where in-sample forecasting is possible under an age-cohort (AC) model. Reference [38] called this approach the "continuous chain ladder" because of its relationship to the chain ladder method used in non-life insurance. Our paper shows that mesothelioma mortality forecasting is one such example even when a period effect is included in the model.

1.3. Aim and Outline

The aim of our paper is to propose, apply and further illustrate in-sample forecasting, by providing a projection of future mesothelioma deaths using simple methodology. We build on the approach of [37] by using the updated dataset from the HSE to conduct a study that forecasts the number of deaths due to mesothelioma. Furthermore, we analyse the differences in forecasts of future expected mesothelioma deaths due in respect of three models: the model applied in this paper, the discrete APC model of [19] and the model using synthetic exposure measures adopted by HSE [39].

Our paper makes two main contributions: First, we illustrate the method in which the approach of [37] addresses the problem of the lack of exposure data by applying the method to a real dataset that would normally pose a challenge to an insurer's claim reserving methodology. Our study expands on the method in terms of usability and applications. While [37] developed an extended continuous chain ladder model which allows for a calendar (period) effect and they illustrated the broad applicability using two empirical examples, we are able to explore further implications of the method when applying it to the challenging problem of mesothelioma mortality forecasting. We show

that the approach is very flexible by introducing a smoothing technique for the temporal effects. Second, we provide an up-to-date estimate of the future number of asbestos related deaths using this new methodology, together with the recent data released by the HSE for 1968–2016.

The remainder of the paper is organised as follows: In Section 2 we formulate the approach of [37] in the context of mortality forecasting and show how it can be applied to provide forecasts of the number of deaths due to mesothelioma. We conduct this in three parts: first, a statistical structured density model is motivated and formulated; second, the density components are estimated non-parametrically; and finally, mortality projections are made into the future. Under the formulated model, we describe how the past data can provide the density component estimates as well as a complete forecast of the future; hence, the name "in-sample forecasting" is used. Section 3 presents the mesothelioma mortality forecasting analysis, together with discussion, before the paper finishes with concluding remarks in Section 4. Further details on the theoretical approach are provided in Appendices A and B.

2. Materials and Methods

2.1. Density Model

Let X denote cohort (birth-year) of an individual and let Y denote age at death for that same individual. Thus, $X + Y$ is the period or calendar year of death. Let us consider X and Y as the two main time effects so we aim to describe the number of deaths as a function of X and Y. According to these variables, past observations $\{(X_i, Y_i); i = 1, \dots, n\}$ are typically supported by a trapezium. Figure 1 shows the special trapezium support of the observations in the dataset analysed later in this paper (see more details in the next section). In this context, the aim is to forecast the number of deaths which will occur in the periods beyond the most recent one where we have observations. The area with observations is a trapezium, and the area we aim to forecast is a triangle. Both areas are presented in Figure 1.
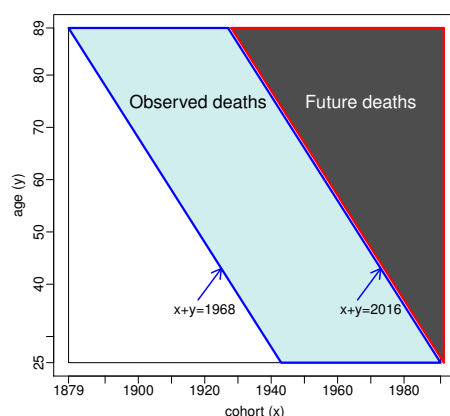

Figure 1. Support of the observations and forecast region.

To derive forecasts for the targeted triangle, Reference [36] suggested modelling and estimating the density of deaths on the full rectangle shown in Figure 1, which is the two dimensional density f of (X, Y). The problem is that, from the available observations, we can only estimate the density from the data contained within the trapezium. The solution to the problem comes from imposing a suitable multiplicative structure for the density. A simple structured density model is as follows:

$$f(x, y) = f_1(x) f_2(y), \tag{1}$$

where f_1 represents the cohort density, and f_2 denotes the age at death density. With the help of a backfitting algorithm, Reference [36] derived estimated components $\hat{f}_1$ and $\hat{f}_2$

from the data in the trapezium and used them to forecast the density on the target triangle. This can be performed since the data provide information for all cohorts and ages involved in the forecasting region.

Model (1) is, however, too simple in many respects. From a mathematical perspective it means that variables X and Y are independent. In the context of this paper, this means that mesothelioma deaths only depend on birth cohort and age of individuals. Reference [36] extended the simple model described above by including a third component f_3 in the multiplicative structure, corresponding to the period or calendar time, that is, the variable $X + Y$. In this paper, we used such an extended model and described the density f of the observations with the following multiplicative structure:

$$f(x, y) = f_1(x) f_2(y) f_3(x + y),\tag{2}$$

where f_3 represents the calendar or period effect on deaths. This model can be described as being a continuous APC model, while model (1) is a continuous AC model.

To simplify the notation in these sections we normalize X_i and Y_i to take values in the unit interval $[0, 1]$ (see left panel of Figure 2) and assume that the observations (X_i, Y_i) are given on the trapezium $\mathcal{I} = \{(x, y) \in [0, 1]^2 : c \leq x + y \leq 1\}$ for some $c > 0$, where $[c, 1]$ is the interval of calendar times where we have observations. The forecast region corresponds to the triangle $\mathcal{I}^{\mathrm{fc}} = \{(x, y) \in [0, 1]^2 : x + y > 1\}$. The assumed trapezium and the forecast region are shown in the right panel of Figure 2.
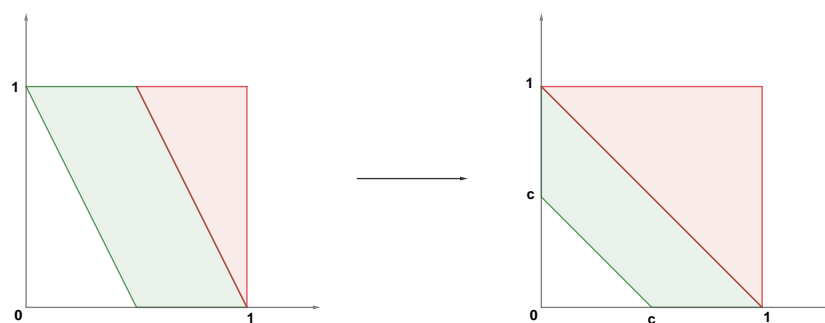

Figure 2. Normalized support of the observations (green) and corresponding forecast region (red).

A key observation at this point is that model (2) is not identified, which means that the functions f_1, f_2 and f_3 are not uniquely determined. Identification of this model relates to the identification of the commonly used age-period-cohort models (see [22,30]). On a logarithmic scale, the model becomes $\log f(x, y) = \log f_1(x) + \log f_2(y) + \log f_3(x + y)$, which can be rewritten as $\log f(x, y) = \log g_1(x) + \log g_2(y) + \log g_3(x + y)$, where the following is the case:

$$\log g_1(x) = -a_1 - bx + \log f_1(x)$$
$$\log g_2(y) = -a_2 - by + \log f_2(y)$$
$$\log g_3(x + y) = a_1 + a_2 + b(x + y) + \log f_3(x + y),$$

with three arbitrary real-valued constants a_1, a_2 and b. This means that the density components (on a logarithmic scale) can only be determined up to two linear trends. To overcome this problem we imposed the following three identification constraints:

$$\int_0^1 f_1(x) dx = 1\tag{3}$$

$$\int_0^1 f_2(y) dy = 1\tag{4}$$

$$f_3(z) = \text{constant} \quad \text{for all } z \in [1 - \kappa, 1],\tag{5}$$

for an appropriate parameter $0 \leq \kappa \leq 1$, which must be determined later on. Notice that determining $\kappa = 1$ means that model (2) becomes the simple continuous AC model (1). In this sense, κ represents the distance from a (continuous) APC model to an AC submodel (the distance being bigger for small values of κ).

Although the conditions (3)–(5) might seem very restrictive, they can always be fulfilled for smooth functions. Conditions (3) and (4) ensure that f_1 and f_2 can be interpreted as proper densities, and they can easily be achieved just by rescaling. Thus, the only assumption to be ensured is that f_3 is constant in the near past (5), which is justified from smoothness considerations. Assuming that $\log f_3(z)$ is differentiable at $z = 1$, it can be approximated by a linear function in that region, i.e., $\log f_3(z) = a + bz + \chi(z)$ for $z \in [1 - \kappa, 1]$, with κ being small, where $\chi(z)$ is approximately zero, and a and b are constants. Notice that this interval might be very small at worst, but it could also be the entire interval $[0, 1]$. Now we can move the linear trend $a + bz$ to $\log(f_1)$ and $\log(f_2)$, and condition (5) is fulfilled. This is illustrated in Appendix A with a simple example.

2.2. Data

We consider data provided by UK Health Service Executive that consist of annual aggregated counts of deaths caused by exposure to asbestos in Great Britain. The original data are given by age levels and calendar year of death between 1968 and 2016 (all given in years). We consider only data corresponding to males with ages between 25 and 89, which gives an array with dimensions 65 (age levels) by 49 (calendar year). The observed total number of deaths (sample size) is $n = 49{,}750$. The authors of [19] analysed the same data at the time but only up to calendar year 2013 by using a discrete APC model. They showed that the main variables to consider in relation to death due to mesothelioma are the cohort and the age (see also [15] for similar conclusions from data up to 2007). Figure 3 shows the data according to these effects and visualizes the special trapezium support available for estimation.
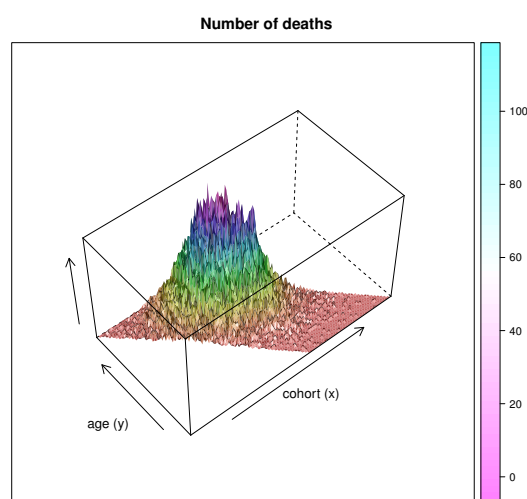

Figure 3. Histogram of the mesothelioma mortality data in UK from 1968 to 2016. The number of deaths are shown according to the cohort (x) and age of death (y), for periods $(x + y)$ between 1968 and 2016.

As discussed in Section 1.2 the only actual data available for forecasting mesothelioma deaths is the observed number of deaths (by age and period of deaths), while the number of people at risk (exposure) is not known. The risk set consists of those who have survived relative to the time of exposure and who have then been exposed. The long latency period of mesothelioma, and its rapid fatal end once discovered, contributes to the problem of finding reliable measures on exposure, as well as data on mortality from competing risks. Many researchers in this area, including the UK Health Service Executive, chose to estimate the exposure and used it in conjunction with the actual data. However constructing these

estimates is nontrivial, and they include a high degree of uncertainty; moreover, regular adjustments are necessary when these measures of exposure are extrapolated [39]. Our approach, which follows that of [15], is to avoid having to estimate the exposure. Instead, we only need to model the observed number of deaths. In fact, in our specific case where we are adopting a continuous approach, we modelled the density of deaths. This suffices since our objective is to forecast aggregated mortality, and its simplicity is an advantage in forecasting.

2.3. Estimation

The density components f_1, f_2 and f_3 in model (2) are the building blocks of any information about mortality in $\mathcal{I}$. If the density $f(x, y)$ is known, all sorts of information could be extracted, such as the number of future deaths in $\mathcal{I}^{\text{fc}}$. Let us denote this number as $D_{\mathcal{I}^{\text{fc}}}$. An estimate can be calculated through the following expression:

$$D_{\mathcal{I}^{\text{fc}}} = \tau \int_{\mathcal{I}^{\text{fc}}} f(x, y) dx dy,$$

where τ is the number of all deaths in the full rectangle in Figure 3. Notice that τ satisfies the relation $\tau \int_{\mathcal{I}} f(x, y) dx dy = D_{\mathcal{I}}$, where $D_{\mathcal{I}}$ is the number of observed deaths in $\mathcal{I}$. This implies that τ can be estimated, in practice, given an estimate $\hat{f}(x, y)$ of $f(x, y)$ evaluated on $\mathcal{I}$ just by computing $\hat{\tau} = D_{\mathcal{I}}(\int_{\mathcal{I}} \hat{f}(x, y) dx dy)^{-1}$. Then, the forecast $D_{\mathcal{I}^{\text{fc}}}$ is obtained by $\hat{D}_{\mathcal{I}^{\text{fc}}} = \hat{\tau} \int_{\mathcal{I}^{\text{fc}}} \hat{f}^{\text{fc}}(x, y) dx dy$, given a forecast $\hat{f}^{\text{fc}}(x, y)$ of $f(x, y)$ on $\mathcal{I}^{\text{fc}}$. Thus, the problem of forecasting the future number of deaths reduces to a density estimation and forecasting problem.

Next, we describe how to estimate the density components f_1, f_2 and f_3 from a data sample $\{(X_i, Y_i) : i = 1, \ldots, n\}$. Consider the following notation: Let $\mathcal{S} = \{(x, y) \in \mathcal{I} : x \leq 1 - \delta, y \leq 1 - \delta\}$, with a small $\delta > 0$, be a subset of $\mathcal{I}$ where we have sufficiently many data points for estimation. Define the following:

$$\mathcal{J}_1(y) = \{x \in [0, 1] : (x, y) \in \mathcal{S}\}$$
$$\mathcal{J}_2(x) = \{y \in [0, 1] : (x, y) \in \mathcal{S}\}$$
$$\mathcal{J}_3(z) = \{x \in [0, 1] : (x, z - x) \in \mathcal{S}\},$$

and the following is the case.

$$f_{w,1}(x) = \int_{\mathcal{J}_2(x)} f(x, y) dy$$

$$f_{w,2}(y) = \int_{\mathcal{J}_1(y)} f(x, y) dx$$

$$f_{w,3}(z) = \int_{\mathcal{J}_3(z)} f(x, z - x) dx.$$

Under model (2), the density components f_1, f_2 and f_3 fulfill the following integral equations:

$$f_1(x) = \frac{f_{w,1}(x)}{\int_{\mathcal{J}_2(x)} f_2(y) f_3(x + y) dy} \tag{6}$$

$$f_2(y) = \frac{f_{w,2}(y)}{\int_{\mathcal{J}_1(y)} f_1(x) f_3(x + y) dx} \tag{7}$$

$$f_3(z) = \frac{\mathbb{1}_{[0, 1 - \kappa]}(z) f_{w,3}(z)}{\int_{\mathcal{J}_3(z)} f_1(x) f_2(z - x) dx} + \frac{\mathbb{1}_{[1 - \kappa, 1]}(z) \int_{1 - \kappa}^{1} f_{w,3}(v) dv}{\int_{1 - \kappa}^{1} \int_{\mathcal{J}_3(v)} f_1(x) f_2(v - x) dx dv}, \tag{8}$$

where $\mathbb{1}_A(x)$ is the indicator function defined by $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise.

In practice, the exact solutions of the above integral equations are unknown because f is unknown. To estimate f_1, f_2 and f_3, we formulated empirical integral equations by substituting f in (6)–(8) by a suitable estimator. We considered the local linear density estimator of [40] with a bandwidth vector (b_1, b_2), which is defined in Appendix B. With this choice, the empirical integral equations are defined as follows:

$$\hat{f}_1(x) = \hat{\phi}_1 \frac{\hat{f}_{w,1}(x)}{\int_{\mathcal{J}_2(x)} \hat{f}_2(y)\hat{f}_3(x+y)\, dy} \tag{9}$$

$$\hat{f}_2(y) = \hat{\phi}_2 \frac{\hat{f}_{w,2}(y)}{\int_{\mathcal{J}_1(y)} \hat{f}_1(x)\hat{f}_3(x+y)\, dx} \tag{10}$$

$$\hat{f}_3(z) = \hat{\phi}_3 \frac{\mathbb{1}_{[0,1-\kappa)}(z)\hat{f}_{w,3}(z)}{\int_{\mathcal{J}_3(z)} \hat{f}_1(x)\hat{f}_2(z-x)dx} + \hat{\phi}_3 \frac{\mathbb{1}_{[1-\kappa,1]}(z)\int_{1-\kappa}^{1} \hat{f}_{w,3}(v)dv}{\int_{1-\kappa}^{1}\int_{\mathcal{J}_3(v)} \hat{f}_1(x)\hat{f}_2(v-x)dxdv} \tag{11}$$

under following constraints:

$$\int_{\mathcal{S}} \hat{f}_1(x)dx = 1, \quad \int_{\mathcal{S}} \hat{f}_2(y)dy = 1 \quad \text{and} \quad \int_{\mathcal{S}} \hat{f}_1(x)\hat{f}_2(y)\hat{f}_3(x+y)dxdy = \hat{\vartheta}, \tag{12}$$

where $\hat{\vartheta} = n^{-1}\sum_{i=1}^{n} \mathbb{1}((X_i, Y_i) \in \mathcal{S})$ is an estimator of $\vartheta = \int_{\mathcal{S}} f(x,y)dxdy$. The coefficients $\hat{\phi}_j$ ($j = 1, 2, 3$) in (9)–(11) are chosen such that the constraints in (12) are satisfied. Here, we have used the notation $\hat{f}_{w,l}$ for the estimator of $f_{w,l}$ above ($l = 1, 2, 3$), which has been obtained by replacing f with $\hat{f}$.

Reference [37] proved the existence of a unique solution for the above empirical equations (see also [41] for related theoretical tools); however, the solution cannot be explicitly obtained, and a backfitting algorithm is required in practice to derive estimates $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$. The algorithm can be written as follows.

Step 0. Let $\hat{f}_1^{[0]}$ and $\hat{f}_2^{[0]}$ be starting values for estimating f_1 and f_2, which satisfy the first two constraints in (12). Calculate the following:

$$\tilde{f}_3^{[0]}(z) = \begin{cases} \dfrac{\hat{f}_{w,3}(z)}{\int_{\mathcal{J}_3(z)} \hat{f}_1^{[0]}(x)\hat{f}_2^{[0]}(z-x)dx} & \text{for } z \in [0, 1-\kappa) \\[4mm] \dfrac{\int_{1-\kappa}^{1} \hat{f}_{w,3}(v)dv}{\int_{1-\kappa}^{1}\int_{\mathcal{J}_3(v)} \hat{f}_1^{[0]}(x)\hat{f}_2^{[0]}(v-x)dxdv} & \text{for } z \in [1-\kappa, 1] \end{cases}$$

and set $\hat{f}_3^{[0]}(z) = \hat{\phi}_3^{[0]}\tilde{f}_3^{[0]}(z)$, where $\hat{\phi}_3^{[0]}$ is chosen such that the third constraint in (12) is satisfied.

Step r. Let $\hat{f}_1^{[r-1]}$, $\hat{f}_2^{[r-1]}$ and $\hat{f}_3^{[r-1]}$ be the backfitting estimates from the previous iteration step. Compute updates as follows.

    (a)   Calculate the following:

$$\tilde{f}_1^{[r]}(x) = \frac{\hat{f}_{w,1}(x)}{\int_{\mathcal{J}_2(x)} \hat{f}_2^{[r-1]}(y)\hat{f}_3^{[r-1]}(x+y)dy}$$

        and set $\hat{f}_1^{[r]}(x) = \hat{\phi}_1^{[r]}\tilde{f}_1^{[r]}(x)$, where $\hat{\phi}_1^{[r]}$ is chosen such that the first constraint of (12) is fulfilled.

    (b)   Calculate the following:

$$\tilde{f}_2^{[r]}(y) = \frac{\hat{f}_{w,2}(y)}{\int_{\mathcal{J}_1(y)} \hat{f}_1^{[r]}(x)\hat{f}_3^{[r-1]}(x+y)dx}$$

and set $\hat{f}_2^{[r]}(y) = \hat{\phi}_2^{[r]} \tilde{f}_2^{[r]}(y)$, where $\hat{\phi}_2^{[r]}$ is chosen such that the second constraint of (12) is satisfied.

(c)   Compute $\hat{f}_3^{[r]}$ analogous to $\hat{f}_3^{[0]}$ in Step 0.

Step r in the algorithm is iterated ($r = 2, 3, \ldots$) until convergence. As a convergence criterion, we evaluate the change $|\hat{f}_i^{[r]}(x) - \hat{f}_i^{[r-1]}(x)| / \max \hat{f}_i^{[r]}(x)$ and stop when it is smaller than a tiny constant $c > 0$, for all $i = 1, 2, 3$. In our empirical analyses, we used $c = 1e - 7$ with a maximum number of iterations of 1000.

Notice that from the above algorithm we obtain estimates for $f_1(x)$ and $f_2(y)$ for $x, y \in [0, 1]$; that is, for all observed cohorts and ages. However we only obtained estimates of $f_3(x + y)$ for $x + y$ in the trapezium $\mathcal{I}$, which is the past period $x + y$. Thus, to derive the required forecasts, we will need to extrapolate f_3. Among other aspects, this issue makes the model (2) more challenging than the simpler model (1). A convenient extrapolation of f_3 is described in Section 2.4.

2.4. Forecasting

Under model (2), from the density components' estimates that were derived in the previous section, we obtained an estimator of the density f of deaths observed in the past calendar times, i.e., in the trapezium $\mathcal{I}$. However, the aim in mesothelioma mortality forecasting is to derive mortality projections for future calendar years, i.e., those lying in the triangle $\mathcal{I}^{fc} = \{(x, y) \in [0, 1]^2 : x + y > 1\}$ shown on the right panel of Figure 2. To satisfy this aim we need to estimate the two-dimensional density f also on $\mathcal{I}^{fc}$. Under model (2), this can be performed from the previous backfitting estimates $\hat{f}_1$ and $\hat{f}_2$, along with a method of extrapolating $\hat{f}_3$ to the future calendar year points ($z = x + y \in (1, 2]$). Recall that the backfitting algorithm only estimates the calendar effect density $f_3(z)$ up to the present time point ($z = 1$).

From the identification constraints (3)–(5), we have assumed that the calendar effect f_3 is constant around the present time point $z = 1$, to be more precise on the interval $[1 - \kappa, 1]$. The authors of [37] used this assumption to extrapolate the calendar effect as a constant into the future, that is, by setting $\hat{f}_3^{fc}(z) = \hat{f}_3(1)$ for $z > 1$ (recall that $z = 1$ represents the more recent observed calendar time). Thus, projections of $f(x, y)$ at future time points $(x, y) \in \mathcal{I}^{fc}$ are given by the following.

$$\hat{f}^{fc}(x, y) = \hat{f}_1(x)\hat{f}_2(y)\hat{f}_3^{fc}(x + y).$$

By contrast to other forecasting strategies, such as the $I(0)$, $I(1)$ and $I(2)$ forecaster of [33], the above extrapolation strategy is a natural method for describing the future based only on the past data (in-sample) and smoothness considerations. While the former forecasters extrapolate the (logarithmic) calendar effect linearly into the future, estimating the slope of such a line in three different ways, our approach eliminates the calendar effect from the model, normalizing it to have zero slope in the recent past by the imposed identification constraints.

At this point there is only one issue left: how to choose, in practice, the constant κ. This parameter can be interpreted as the length of the recent past which should be used to estimate and forecast the calendar effect. Therefore, it is of interest in practice to illustrate the effect of such a parameter on the forecasts. In our application to mesothelioma mortality, we perform this and conclude with a data-driven choice derived by cross-validation. The cross-validation method estimate κ from the data as follows: pick some small $\delta > 0$ and define the following.

$$\mathcal{S}_\delta^< = \{(x, y) \in \mathcal{S} : x + y \leq 1 - \delta\}$$
$$\mathcal{S}_\delta^> = \{(x, y) \in \mathcal{S} : x + y > 1 - \delta, x \leq 1 - \delta, y \leq 1 - \delta\}.$$

Let $\mathcal{D}$ be the set of data points (X_i, Y_i) that lie in $\mathcal{S}_\delta^<$, that is, $(X_i, Y_i) \in \mathcal{S}_\delta^<$. For any $\kappa \in (\delta, 1]$, compute the backfitting estimators $\hat{f}_1^\kappa, \hat{f}_2^\kappa, \hat{f}_3^\kappa$ from the data sample $\mathcal{D}$, as described in Section 2.3, where the set $\mathcal{S}$ is replaced by $\mathcal{S}_\delta^<$. Next, we define the following.

$$\hat{f}_3^\kappa(z) = \hat{f}_3^\kappa(1-\delta) \quad \text{for } z \in (1-\delta, 1],$$

The estimator of κ is defined as follows:

$$\hat{\kappa} = \arg\min_{\kappa \in (\delta, 1]} \mathrm{CV}(\kappa). \tag{13}$$

where the following is the case.

$$\mathrm{CV}(\kappa) = \int_{\mathcal{S}_\delta^>} \{\hat{f}^\kappa(x, y)\}^2 dx dy - \frac{2}{n} \sum_{i=1}^{n} \mathbb{1}((X_i, Y_i) \in \mathcal{S}_\delta^>) \hat{f}^\kappa(X_i, Y_i).$$

The justification of the above criterion comes from the fact that CV is an estimator of the Mean Integrated Squared Error (MISE) of $\hat{f}^\kappa(x, y)$ in the set $\mathcal{S}_\delta^>$, which ideally one would minimise to choose κ. To observe this, we expand the MISE as follows.

$$\mathrm{MISE}(\hat{f}^\kappa(x,y)) = \int_{\mathcal{S}_\delta^>} (\hat{f}^\kappa(x,y) - f(x,y))^2 dx dy$$
$$= \int_{\mathcal{S}_\delta^>} \hat{f}^\kappa(x,y)^2 - 2f(x,y)\hat{f}^\kappa(x,y) + f(x,y)^2 dx dy.$$

Since the last term is positive and does not dependent on κ, we can minimise the above expression for MISE and ignore the final term. Then, as the second term depends on the unknown f, we can replace it with the simple non-parametric estimator $n^{-1} \sum_{i=1}^{n} \mathbb{1}((X_i, Y_i) \in \mathcal{S}_\delta^>)$. This provides the above expression for $\mathrm{CV}(\kappa)$.

3. Results

Using the methodology described above, we have analysed a dataset consisting of annual aggregated counts of deaths caused by exposure to asbestos in Great Britain for males aged between 25 and 89. From this data, we can update the results from the previous analysis of [19] using the additional data and applying model (2), which allows us to take into account the calendar year effect on the mesothelioma deaths.

Therefore, we assume that the two-dimensional density $f(x,y)$, with x denoting the cohort and y the age, can be written as the product of three functions: the densities corresponding to the cohort and the age effects, f_1 and f_2, as well as a third function, f_3, describing the effect of the year (period) of death. In order to estimate the three density components, we have used the backfitting algorithm described in Section 2.3. For the local linear estimator $\hat{f}$ of the two-dimensional density f, we have considered bandwidths $\hat{b}_1 = 6$, $\hat{b}_2 = 4.2$ years. The bandwidths $\hat{b}_1$ and $\hat{b}_2$ were obtained as follows: we first computed the common cross-validated bandwidths $\tilde{b}_1$ and $\tilde{b}_2$ for the local linear estimator $\hat{f}$ and then rescaled them by the factor $n^{-1/5}/n^{-1/6}$. The justification of this rescaling is based on theory. The authors of [37] proved the consistency of the backfitting estimates assuming that the component bandwidths satisfy the condition $n^{1/5}b_j \to c_j$ for some constant $c_j > 0$ ($j = 1, 2$), i.e., they have convergence order of $n^{-1/5}$. The cross-validated estimates derived for the two-dimensional density have order $n^{-1/6}$ (see [40]), so rescaling the cross-validated bandwidths with the above factor provides the theoretical requirements.

The estimated density components are shown in Figure 4. The graphs display the estimates produced by the backfitting algorithm for different values of the parameter κ. Recall that this parameter defines the length of the most recent time interval where the calendar effect function f_3 is constant. It can be also interpreted as the length of the recent past, which is used to estimate the calendar effect. For an easier interpretation here, we

provide this parameter in number of years so the value $\kappa = 49$ corresponds to a constant calendar effect over the whole period from 1968 to 2016, $\kappa = 10$ corresponds to a constant calendar effect in the last 10 years, i.e., from 2007 to 2016, and non-constant from 1968 to 2006; in general, $\kappa = 49 - p$ corresponds to a constant calendar effect in the last p years, i.e., from $2016 - p + 1$ to 2016, and non-constant from 1968 to $2016 - p$. This can be seen in the last graph of Figure 4 where the bigger κs correspond to nearly constant estimates of f_3, while smaller values allow for general shapes. The estimates of f_1 and f_2 also vary slightly with the value of κ. Notice that the shapes of the densities do not allow us to observe the variations in more detail in the graphs.
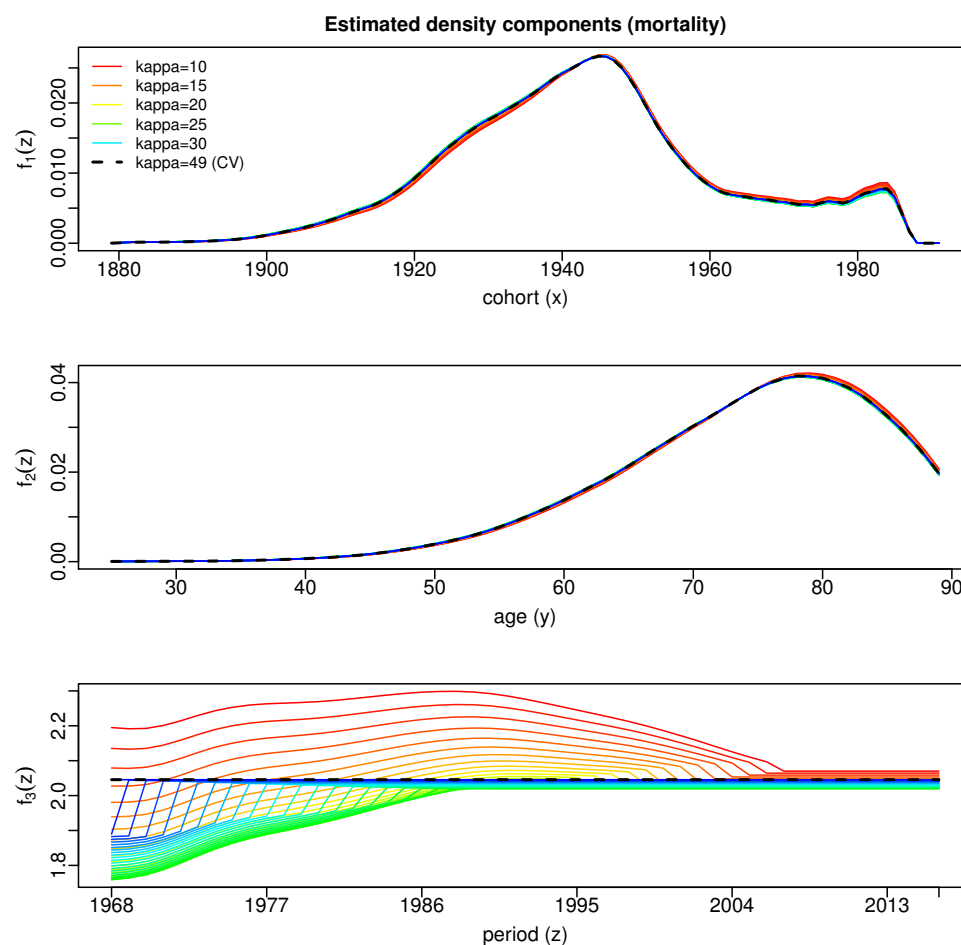

Figure 4. Estimated density components for the mesothelioma mortality data considering different κ values (given in years). The dashed black line shows the estimated density components for the $\kappa = 49$ years chosen by cross-validation.

Applying the forecasting method described in Section 2.4 in conjunction with the density component estimates for different κ's, we have calculated the predicted total number of deaths in future years (that is, in the years following 2016). The results are shown in Figure 5 for $\kappa = 10, \ldots, 49$, along with the observed number of past deaths. We can observe that κ does not have a big effect on the forecasts, which agrees with the slight variations in the density component estimates for each κ shown in Figure 4.

It is of interest to predict the peak value (highest number of deaths) as well as the year of peaking. From Figure 5, we can observe that the peak had already been reached in 2016, and it is confirmed for all the κ values considered. The peak value was 2101 observed deaths. This agrees with the provisional data for 2017 published by the HSE [39], where the number of deaths in the year 2017 had decreased to 2087.

To derive our final forecasts, we consider the cross-validation method defined above to choose the value of κ, see Equation (13). In order to minimize the function CV, we evaluate it on $\kappa = 10, \ldots, 49$. The CV function is quite flat, showing slightly smaller values for the bigger κ's. This seems to be in line with the stable behaviour of the density component estimates shown in Figure 4 and suggests choosing the maximum value of κ, which is $\kappa = 49$ years. This value corresponds to the case of f_3 being constant for every period, that is, it corresponds to a simple age-cohort model. For this choice we predicted 2063 deaths in the year 2017, which is a bit lower than the available provisional figure of 2087 for 2017 [39]. The number of deaths decreases slowly until 2032 deaths were reached in 2020. Table 1 shows our forecasts until 2022 using the full dataset (third column), as well as using only data up to 2013 (sixth column). The available data are shown for assessing forecasts derived from data up to 2013 (second column).
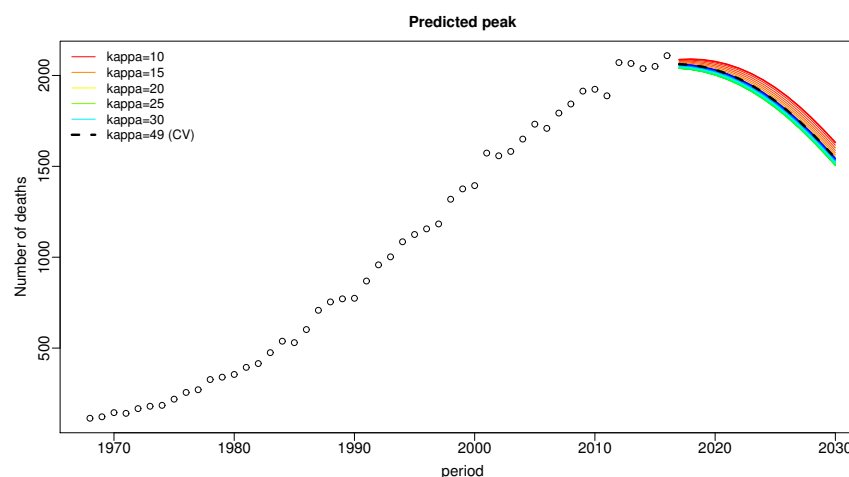

Figure 5. Forecasts of the annual number of deaths using different values of κ. The observed past numbers of deaths are indicated by dots.

Table 1. Mesotheolioma mortality forecasts in the UK. Five forecasts of numbers of deaths are compared: "our-201x" corresponds to our proposal using data up to 201x and constant calendar effect; "apc-201x*" to the discrete approach of [19] using data up to 201x and truncating cohorts from 1966 (x = 3, 6); and HSE projections also using provisional data for 2017, "HSE-2017p". Available data are shown for assessing forecasts.

Period	Data	Our-2016	apc-2016*	HSE-2017p	Our-2013	apc-2013*
2014	2032				2048	2056
2015	2042				2062	2070
2016	2101				2071	2077
2017	2087	2063	2069		2074	2079
2018		2058	2062	2068	2072	2074
2019		2048	2049	2036	2063	2063
2020		2032	2030	1994	2049	2045
2021		2010	2002	1943	2028	2018
2022		1982	1969	1885	2002	1988

For comparison purposes we have included in the table forecasts derived from the discrete approach of [19] (fourth and last columns) and the more recent HSE forecasts [39] (fifth column). The discrete approach of [19] is computed by truncating the data corresponding to the youngest cohorts, that is, from 1966 as those authors suggested. This is necessary since the approach does not perform any smoothing and cannot properly deal with sparsity in the data corresponding to those cohorts. These forecasts have been computed using the `apc` package [42] and are shown for years up to 2022. Our approach and the discrete approach of [19] provide similar forecasts, but we do not need to perform

any arbitrary truncation in contrast to the discrete approach. Moreover, truncating cohorts from 1966 means that mortality is only projected from ages above 50, which might explain the slightly lower forecasts from the discrete model. On the other hand, HSE provides a similar forecast for calendar year 2018 but differs substantially for future periods. Figure 6 shows the differences in the shapes of forecasts being compared. HSE projects a notably faster decline of deaths than the other approaches. For reference, we have also shown in this figure the forecasts published in [19] using (truncated) data up to 2013 (the truncation in this case resulted in projections only for ages above 47 years). Moreover, we have added our forecasts with data up to 2013 (without truncation) and constant calendar effect (corresponding to the value of κ chosen by cross-validation).
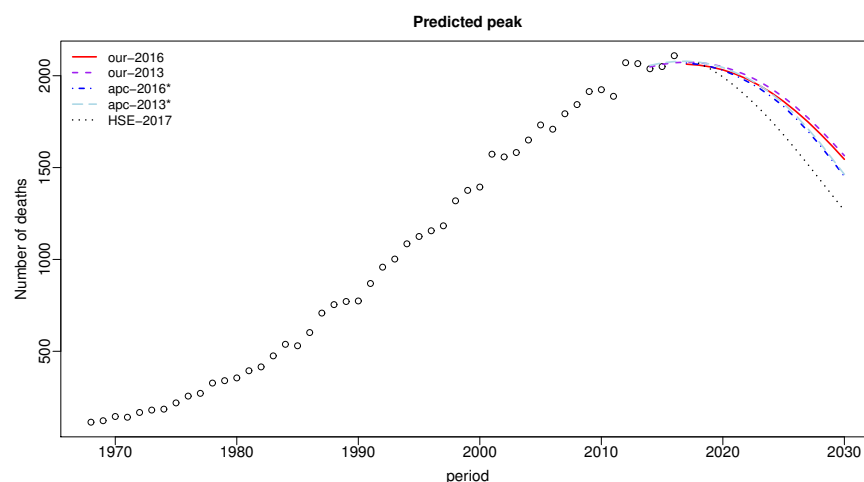

Figure 6. Comparison of forecasts. The observed past numbers of deaths are indicated by dots.

By restricting the data for estimation up to 2013, we can assess whether our proposal would have been able to predict the peak in 2016. Figure 7 shows the peak forecasts for different values of κ along with the year of peak. Peak values vary with κ between 2032 and 2077 and the year of peaking between 2016 and 2018. The observed peak of 2101 in 2016 is, therefore, hardly predicted using data up to 2013. The same happens with the forecasts derived by [19], which predicted a peak of 2079 deaths in 2017. This seems to be natural since the statistical projections describe the expected future mortality as a smooth curve, while the actual numbers in years close to the peak are expected to fluctuate above and below due to year-on-year random variation [39].
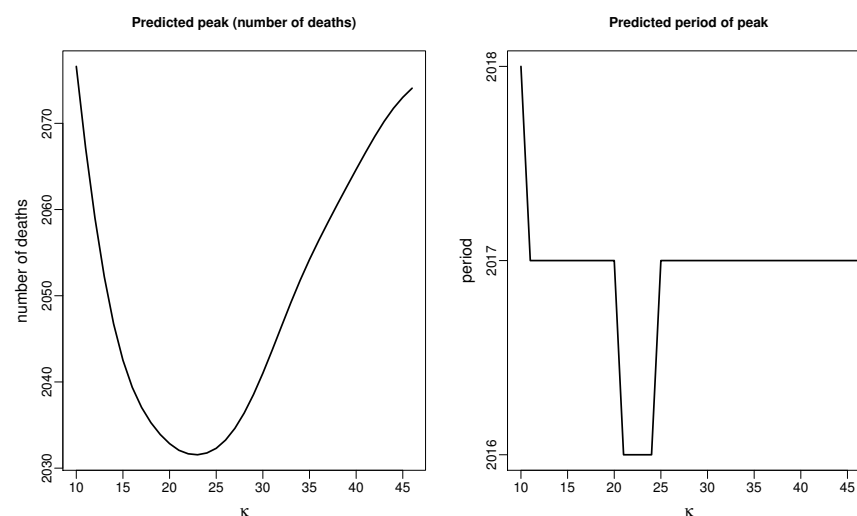

Figure 7. Peak forecasts (**left**) and year of peak (**right**) for different values of κ. Based on data up to 2013.

4. Discussion and Conclusions

Standard methods and common benchmarks in the literature on mortality forecasting rely on dose–response models where both deaths and exposure are observed. Such methods involve the non-trivial and risky exercise of estimating the exposure when it is not known, as happens in the case of asbestos-related mortality. Our paper demonstrates that the methodology of [37] can be another benchmark with important benefits. First, it does not require any modelling, estimating and extrapolating of the exposure when it is unknown. This makes the approach more robust compared with those that are more detailed in this regard, especially when the model for the exposure is mis-specified. Second, it is very intuitive due to its connection with the popular APC models. Third, it takes advantage of the powerful non-parametric structured models, which exhibit excellent theoretical and practical properties, compared to the standard (discrete) APC models. Finally, forecasting is entirely determined by the data, avoiding the need to use time series modelling and other more sophisticated extrapolation techniques. This further contributes to the robustness of the approach in practice.

We applied our method to actual data consisting of the number of deaths due to mesothelioma between 1968 and 2016, which we obtained from the Health and Safety Executive. From this, we have been able to produce an up to date forecast for the number of deaths in the future. We have also compared the results from our model to those from the discrete AC model of [19] and the model used by the Health and Safety Executive. While our forecasts, and those from the discrete age-cohort model, show very similar shapes, the HSE forecasts differ notably, showing a much faster decline in the number of deaths up to calendar year 2030. In addition, we have provided forecasts with truncated data which could be compared to data for the following years. In all cases, the forecasts were reasonable but differed depending on the constant κ, which is the length of the interval in the past for which the period function is set as constant. While choosing κ is challenging, we explained a cross validation approach on how to choose it in practice.

Modelling the number of deaths by our density approach suffices when the objective is to forecast aggregated mortality. However, more detailed information might be required beyond this objective. For instance, if a further study considers the prevention of deaths due to mesothelioma, then it would be necessary to model the length of the latency period and to take into account survival from competing risks.

A limitation of previous approaches compared in this paper ([19,39]) is that the age-profile is assumed to be common for all cohorts [12]. The model considered in this paper relaxes this assumption by introducing some dependence between age and cohorts. More general structures could be assumed under a similar density approach, but with the risk of increasing the uncertainty arising out of such more general models. Therefore, further research would be required to find a good compromise between model complexity and uncertainty. Evaluating the uncertainty of the forecasts is another issue which will require further work, e.g., to derive confidence bands for the forecasts.

Appendix A

Here, we describe how the identification constraint (5) can be fulfilled in practice by using a simple example. Recall that, under smoothness assumptions, this amounts to removing from f_3 a log-linear trend in the interval $[1 - \kappa, 1]$ so it becomes (approximately) constant.

Define the density components in model (2) as $f_1(x) = 1$ (cohort effect), $f_2(y) = \exp(-y)$ (age effect), for $0 \leq x, y \leq 1$ and $f_3(z) = \exp(z/3)$, for $0 \leq z \leq \kappa$, and $f_3(z) = \exp(-\kappa/2 + 5z/6)$ and for $1 - \kappa \leq z \leq 1$ with $\kappa = 0.4$ (period effect). The three components are displayed in black in Figure A1. Now, we remove the log-linear trend of f_3 in the interval $[0.6, 1]$, that is, $\exp(5z/6) = \exp(5x/6)\exp(5y/6)$, and allocate $\exp(5x/6)$ into $f_1(x)$ and $\exp(5y/6)$ into $f_2(y)$. The resulting density components are displayed in red in the same figure. Thus, we have rewritten the density components in the model in such a manner that f_3 fulfills the constraint (5). Notice that this practice can be extended to any function f_3 as long it is smooth around $z = 1$ for at least a small κ.
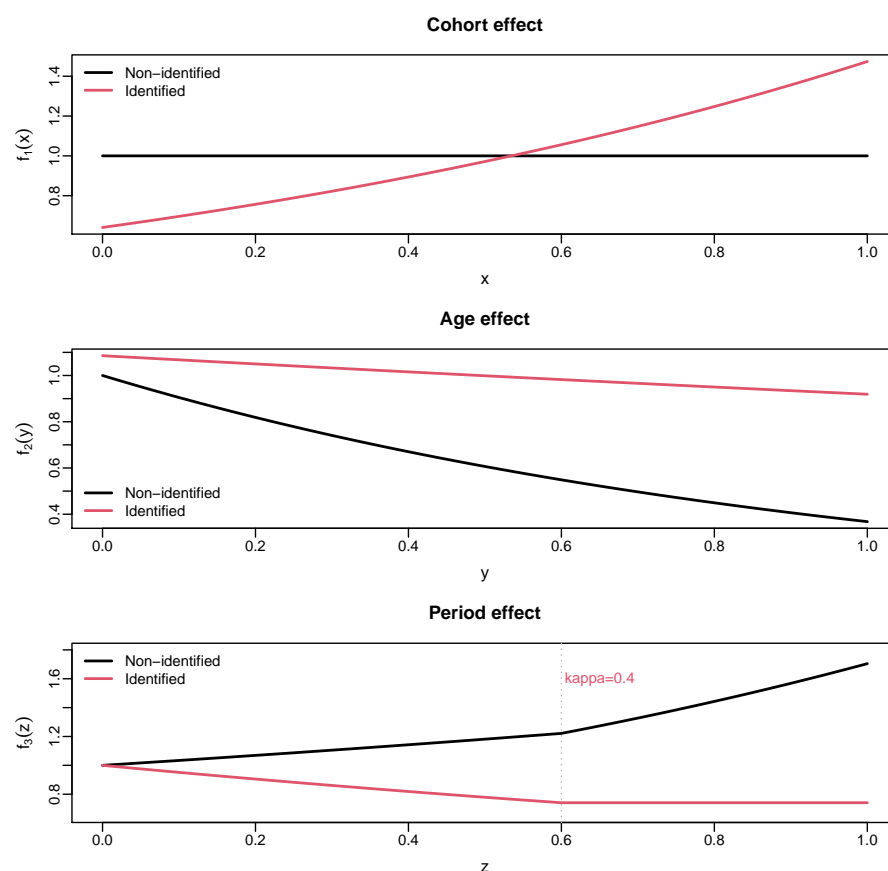

Figure A1. Illustration of how the identification works in an example. The log-linear trend of f_3 in the interval $[1 - \kappa, 1]$ is removed and allocated into the components f_1 and f_2.

Appendix B. Two-Dimensional Local Linear Density Estimator

The local linear estimator of a two-dimensional density function $f(x, y)$ was introduced by [40] as follows. Let the following:

$$\tilde{f}_{h_1,h_2}(x,y) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} K\left(\frac{X_i - x}{b_1}\right) K\left(\frac{Y_i - y}{b_2}\right) W_i$$

be a standard kernel density estimator of f (see for example [43]), where $W_i = 1((X_i, Y_i) \in \mathcal{S})$, K is a two-dimensional kernel function, and (h_1, h_2) is the bandwidth vector. Consider the following minimization problem:

$$\hat{\boldsymbol{\eta}}(x,y) = \arg\min_{\boldsymbol{\eta}=(\eta_0,\eta_1,\eta_2)} \lim_{h_1,h_2 \to 0} \int_{\mathcal{S}} \left[\tilde{f}_{h_1,h_2}(v,w) - \boldsymbol{a}(v,w;x,y)^{\top} \boldsymbol{\eta}(x,y)\right]^2$$
$$\times K\left(\frac{v-x}{b_1}\right) K\left(\frac{w-y}{b_2}\right) dv dw, \tag{A1}$$

where $\boldsymbol{a}(v,w;x,y) = (1, (v-x)/b_1, (w-y)/b_2)^{\top}$, and let $\hat{\boldsymbol{\eta}} = (\hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2)$ denote its solution. It can be shown [37] that the following is the case:

$$\hat{\boldsymbol{\eta}}(x,y) = \boldsymbol{A}(x,y)^{-1}\boldsymbol{b}(x,y),$$

where the following results.

$$\boldsymbol{A}(x,y) = \int_{\mathcal{S}} \boldsymbol{a}(v,w;x,y)\boldsymbol{a}(v,w;x,y)^{\top} b_1^{-1} b_2^{-1} K\left(\frac{v-x}{b_1}\right) K\left(\frac{w-y}{b_2}\right) dv dw$$

$$\boldsymbol{b}(x,y) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{a}(X_i, Y_i; x, y) h_1^{-1} h_2^{-1} K\left(\frac{X_i - x}{b_1}\right) K\left(\frac{Y_i - y}{b_2}\right) W_i.$$

The local linear estimator $\hat{f}$ is defined as the first component of the vector $\hat{\boldsymbol{\eta}}$.

References

1. Selby, K. Mesothelioma Statistics. Available online: https://www.asbestos.com/mesothelioma/statistics/ (accessed on 25 August 2021).
2. O'Reilly, K.M.; Mclaughlin, A.M.; Beckett, W.S.; Sime, P.J. Asbestos-related lung disease. *Am. Fam. Physician* **2007**, *75*, 683–688.
3. UK Asbestos Working Party. Update from UK Asbestos Working Party. Available online: www.actuaries.org.uk/practice-areas/general-insurance/research-working-parties/uk-asbestos (accessed on 18 December 2020).
4. AM Best. Asbestos and Environmental Losses Continue. Available online: http://news.ambest.com/articlecontent.aspx?refnum=281133&altsrc=43 (accessed on 18 December 2020).
5. Janssen, F. Advances in mortality forecasting: Introduction. *Genus* **2018**, *74*, 21. [CrossRef]
6. Lee, R.; Carter, L. Modeling and forecasting U.S. mortality. *J. Am. Stat. Assoc.* **1992**, *87*, 659–671. [CrossRef]
7. Hatzopoulos, P.; Haberman, S. A parameterized approach to modeling and forecasting mortality. *Insur. Math. Econ.* **2009**, *44*, 103–123. [CrossRef]
8. Booth, H.; Tickle, L. Mortality modelling and forecasting: A review of methods. *Ann. Actuar. Sci.* **2008**, *3*, 3–43. [CrossRef]
9. Hyndman, R.J.; Ullah, M. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* **2007**, *51*, 4942–4956. [CrossRef]
10. Renshaw, A.; Haberman, S. A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insur. Math. Econ.* **2006**, *38*, 556–570. [CrossRef]
11. Russolillo, M.; Giordano, G.; Haberman, S. Extending the Lee-Carter model: a three-way decomposition. *Scand. Actuar. J.* **2011**, *2011*, 97–117. [CrossRef]
12. Hodgson, J.; McElvenny, D.; Darnton, A.; Price, M.; Peto, J. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *Br. J. Cancer* **2005**, *92*, 587–593. [CrossRef]
13. Tan, E.; Warren, N. Projection of mesothelioma mortality in Great Britain. In *Health and Safety Executive, Research Report*; HSE Books: Norwich, UK, 2009; p. 728.
14. Tan, E.; Warren, N.; Darnton, A.J.; Hodgson, J.T. Projection of mesothelioma mortality in Britain using Bayesian methods. *Br. J. Cancer* **2010**, *103*, 430–436. [CrossRef] [PubMed]
15. Miranda, M.D.M.; Nielsen, B.; Nielsen, J.P. Inference and forecasting in the age–period–cohort model with unknown exposure with an application to mesothelioma mortality. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2015**, *178*, 29–55. [CrossRef]

16. Trotta, A.; Santana, V.S.; Andreozzi, L. P010 Forecasting of Mesothelioma Mortality in Argentina, 2014–2023. Available online: http://dx.doi.org/10.1136/oemed-2016-103951.335 (accessed on 24 August 2021).
17. Oddone, E.; Bollon, J.; Nava, C.R.; Consonni, D.; Marinaccio, A.; Magnani, C.; Gasparrini, A.; Barone-Adesi, F. Effect of Asbestos Consumption on Malignant Pleural Mesothelioma in Italy: Forecasts of Mortality up to 2040. *Cancers* **2021**, *13*, 3338. [CrossRef]
18. Algranti, E.; Saito, C.A.; Carneiro, A.P.S.; Moreira, B.; Mendonça, E.M.C.; Bussacos, M.A. The next mesothelioma wave: mortality trends and forecast to 2030 in Brazil. *Cancer Epidemiol.* **2015**, *39*, 687–692. [CrossRef]
19. Martínez-Miranda, M.D.; Nielsen, B.; Nielsen, J.P. Simple benchmark for mesothelioma projection for Great Britain. *Occup. Environ. Med.* **2016**, *73*, 561–563. [CrossRef] [PubMed]
20. Zehnwirth, B. *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital*; Casualty Actuarial Society Forum: Arlington, VA, USA, 1994; Volume 2, pp. 447–606.
21. England, P.D.; Verrall, R.J. Stochastic claims reserving in general insurance. *Br. Actuar. J.* **2002**, *8*, 443–518. [CrossRef]
22. Kuang, D.; Nielsen, B.; Nielsen, J.P. Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **2008**, *95*, 979–986. [CrossRef]
23. O'Brien, R. *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data*; Chapman and Hall CRC Press: Boca Raton, FL, USA, 2014.
24. Smith, T.R.; Wakefield, J. A review and comparison of age-period-cohort models for cancer incidence. *Stat. Sci.* **2016**, *31*, 591–610. [CrossRef]
25. Carstensen, B. Age–period–cohort models for the Lexis diagram. *Stat. Med.* **2007**, *26*, 3018–3045. [CrossRef]
26. Clayton, D.; Schifflers, E. Models for temporal variation in cancer rates. I: Age–period and age–cohort models. *Stat. Med.* **1987**, *6*, 449–467. [CrossRef]
27. Clayton, D.; Schifflers, E. Models for temporal variation in cancer rates. II: Age–period–cohort models. *Stat. Med.* **1987**, *6*, 469–481. [CrossRef] [PubMed]
28. Keiding, N. Statistical inference in the Lexis diagram. *Philos. Trans. R. Soc. Lond. Ser. Phys. Eng. Sci.* **1990**, *332*, 487–509.
29. Beutner, E.A.; Reese, S.; Urbain, J.P. Identifiability issues of age-period and age-period-cohort models of the Lee-Carter type. *Insur. Math. Econ.* **2017**, *75*, 117–125. [CrossRef]
30. Nielsen, B.; Nielsen, J.P. Identification and forecasting in mortality models. *Sci. World J.* **2014**, *2014*, 347043. [CrossRef] [PubMed]
31. Berzuini, C.; Clayton, D. Bayesian analysis of survival on multiple time scales. *Stat. Med.* **1994**, *13*, 823–838. [CrossRef]
32. Yang, Y.; Land, K.C. Age–period–cohort analysis of repeated cross-section surveys: Fixed or random effects? *Sociol. Methods Res.* **2008**, *36*, 297–326. [CrossRef]
33. Kuang, D.; Nielsen, B.; Perch Nielsen, J. Forecasting in an extended chain-ladder-type model. *J. Risk Insur.* **2011**, *78*, 345–359. [CrossRef]
34. Hunt, A.; Blake, D. Identifiability in age/period/cohort mortality models. *Ann. Actuar. Sci.* **2020**, *14*, 500–536. [CrossRef]
35. Lee, Y.; Mammen, E.; Nielsen, J.P.; Park, B. Asymptotics for in-sample density forecasting. *Ann. Stat.* **2015**, *43*, 620–651. [CrossRef]
36. Mammen, E.; Miranda, M.D.M.; Nielsen, J.P. In-sample forecasting applied to reserving and mesothelioma mortality. *Insur. Math. Econ.* **2015**, *61*, 76–86. [CrossRef]
37. Mammen, E.; Martínez-Miranda, M.D.; Nielsen, J.P.; Vogt, M. Calendar effect and in-sample forecasting. *Insur. Math. Econ.* **2021**, *96*, 31–52. [CrossRef]
38. Miranda, M.D.M.; Nielsen, J.P.; Sperlich, S.; Verrall, R. Continuous Chain Ladder: Reformulating and generalizing a classical insurance problem. *Expert Syst. Appl.* **2013**, *40*, 5588–5603. [CrossRef]
39. Health and Safety Executive (HSE). Mesothelioma Statistics for Great Britain. Annual Statistics. 2019. Available online: www.hse.gov.uk/statistics/ (accessed on 14 July 2021).
40. Neilsen, J.P. Multivariate boundary kernels from local linear estimation. *Scand. Actuar. J.* **1999**, *1999*, 93–95. [CrossRef]
41. Mohammadi, B.; Shole Haghighi, A.A.; Khorshidi, M.; De la Sen, M.; Parvaneh, V. Existence of Solutions for a System of Integral Equations Using a Generalization of Darbo's Fixed Point Theorem. *Mathematics* **2020**, *8*, 492. [CrossRef]
42. Nielsen, B. apc: Age-Period-Cohort Analysis, R Package Version 1.4. Available online: https://cran.r-project.org/web/packages/apc/index.html (accessed on 14 July 2021).
43. Wand, M.; Jones, M. *Kernel Smoothing*; Chapman and Hall: London, UK, 1995.