

MORTALITY SHOCKS & REPORTING DELAYS IN PORTFOLIO DATA

By S. J. Richards



Mortality shocks and reporting delays in portfolio data

Richards, Stephen J.*

February 8, 2021

Abstract

The covid-19 pandemic requires that actuaries track short-term mortality fluctuations in the portfolios they manage. This demands methods that not only operate over much shorter time periods than a year, but that also deal with reporting delays. In this paper we consider both semi-parametric and fully parametric approaches for tracking portfolio mortality levels in continuous time. We identify both seasonal patterns and mortality shocks, thus providing a benchmark for covid-19 in terms of a portfolio's own past experience. A model is presented that allows for seasonal variation and reporting delays. We find that an estimate of mortality reporting delays can be made from a single extract of experience data and used to forecast unreported deaths. Results are given for annuity portfolios in France, the UK and the USA.

Keywords: covid-19, seasonal variation, OBNR, IBNR.

1 Introduction

The covid-19 pandemic [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020] created the need to measure and track mortality shocks in portfolios managed by actuaries. The intense nature of the covid-19 mortality shock in many countries means that traditional methods based around annual q_x -style mortality rates are inadequate: mixing periods of shock and non-shock mortality will under-state the true intensity of a mortality spike. Furthermore, risk managers and decision-makers cannot wait until a full year's experience becomes available, while administration functions need to plan staffing levels from week to week for processing surges in death notifications. Something akin to real-time mortality reporting is therefore needed, so continuous-time methodologies are strongly preferred.

There are two related practical issues. First, the short-term nature of the covid-19 shock in some territories carries the risk of confounding with normal seasonal variation; see Rau [2007] for a comprehensive overview of human seasonal mortality and Richards et al. [2020] for the specifics of seasonal mortality among pensioners and annuitants. Any assessment of the impact of covid-19 on an insured portfolio must therefore take into account typical seasonal variation. Second, the most up-to-date available data will be affected by unreported deaths, so any methodology must handle the inherent delays in death reporting.

This paper presents semi-parametric and fully parametric approaches to tracking portfolio mortality in continuous time, and considers their abilities to address these issues. We present a simple, semi-parametric estimator that tracks portfolio mortality levels by operating on intervals defined by the gaps between dates of death. This means daily estimation of portfolio mortality levels where the data support it. We illustrate our approach by applying the estimator to annuitant portfolios in three countries, revealing the timing and impact of the covid-19 shock and benchmarking it against past

*stephen@longevity.co.uk, Longevity Ltd, Edinburgh. www.longevity.co.uk. <http://orcid.org/0000-0001-6859-6347>

seasonal variation in those same portfolios. We further present a fully parametric model that allows for delays in death reporting, and thus provides continuous forecasts of emerging portfolio mortality.

The plan of the rest of this paper is as follows: Section 2 describes the data sets used in the paper, while Section 3 defines and illustrates a semi-parametric approach. Section 4 considers delays in reporting and recording of deaths. Section 5 presents a parametric methodology for dealing with such delays and Section 6 considers its effectiveness in terms of fit; Section 7 examines the ability of the parametric OBNR (occurred-but-not-reported) term to forecast delayed mortality, and how to use it to improve the semi-parametric estimator of Section 3; Section 8 concludes.

2 Data description

The data sets used in this paper comprise individual records from the insurer annuity portfolios described in Table 1. Due to the financial interest in not paying longer than necessary, administrators maintain accurate records of when annuities commence and when they cease. Such portfolios are like longitudinal studies with continuous recruitment: new benefit records are set up as people retire, but also when the death of the primary annuitant triggers an annuity to a surviving spouse.

Table 1: Overview of portfolios and data extracts available. The naming convention was chosen for consistency with Richards et al. [2020], as some of the same portfolios are used.

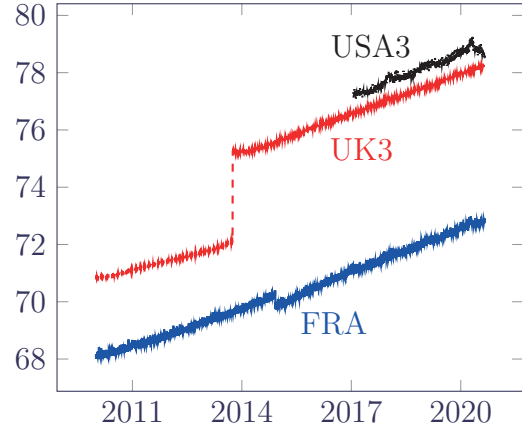
Portfolio	Description	Date of extract	Cumul. total deaths	Covid-19 pandemic included	In-force on 1 April 2020
FRA	Insurer portfolio of voluntary top-up pensions for employees of higher-education institutions around France.	June 2019	42,284	No	n/a
		June 2020	45,373	Yes	251,528
		Sept. 2020	47,026	Yes	251,330
UK3	Individual annuitants from defined-contribution personal pensions, widely spread around the UK.	June 2020	108,003	Yes	146,346
		Sept. 2020	109,878	Yes	146,269
USA3	Annuitants from bulk buy-outs spread around the USA.	Sept. 2020	145,153	Yes	723,762

Direct extracts were made from the administration systems for the FRA and UK3 portfolios, as recommended by Macdonald et al. [2018, Section 2.2] to get the most up-to-date data. Extracts were taken at different times, and the corresponding cumulative death counts are shown in Table 1. The most recent extract will be used for estimating the impact of the covid-19 mortality shock in Section 3, while earlier extracts for the same portfolio will be used for assessing delays in death reporting and recording in Section 4.

Policy records were validated using the checks described in Macdonald et al. [2018, Sections 2.3 and 2.4]. Insurer annuity portfolios often contain two or more policies per person, and so a data-preparation stage of deduplication is normally required for statistical modelling; see Macdonald et al. [2018, Section 2.5] for discussion of various approaches. The UK3 portfolio in particular contains many such duplicates. However, the FRA portfolio did not contain enough detail in one of the extracts to permit deduplication, so we will use the policy-level records without deduplication in this paper. This means that our mortality data are over-dispersed [Djeundje and Currie, 2011] due to some individuals receiving more than one annuity.

Figure 1 shows that the average age increased over the exposure period in all three portfolios, indicating that the numbers of new policies are not enough to stop the portfolios ageing. The large step increase in average age for UK3 is due to the transfer out of a block of annuities, while the smaller step decrease in average age for FRA is due to 9,572 new annuities in December 2014. We will therefore restrict our attention to experience data on or after 1st January 2015 to avoid step changes in the age composition of portfolios.

Figure 1: Average age over time of in-force annuitants in Table 1.



3 A semi-parametric approach

The mortality hazard, $\mu_{x,y}$, is a function of age, x , and calendar time, y , plus other risk factors such as gender, pension size and others [Richards et al., 2013]. Various non-parametric estimators of mortality functions exist, but they typically only have one time-varying element (usually age). For example, the Nelson-Aalen estimator [Nelson, 1958] of the integrated mortality hazard and the Kaplan-Meier estimator of the survivor function [Kaplan and Meier, 1958] are usually defined with respect to age, leaving unmodelled heterogeneity with respect to calendar time. In this section we consider non-parametric estimators defined with respect to calendar time, leaving age and other factors as unmodelled sources of heterogeneity. We note that such estimators will be in no sense useful for estimating the mortality of a portfolio for actuarial purposes like pricing and reserving. However, over short periods of time most portfolios do not change much in composition, and so such estimators might reveal insights into short-term variation in mortality levels.

Adapting the notation of Macdonald et al. [2018, page 140], we consider a Nelson-Aalen estimator, $\hat{\Lambda}_{y,t}$, of the cumulative hazard from calendar time y to $y+t$. We turn dates into real numbers by taking the year and adding the number of days elapsed divided by the number of days in that year. For example, 2020 is a leap year, so the date 2020-03-14 is represented as 2020.199454 ($=2020 + 73/366$). We create a set $\{y+t_i\}$ of n_y distinct times (dates) of death. We define d_{y+t_i} as the number of deaths occurring at time $y+t_i$ and define $l_{y+t_i^-}$ as the number of lives immediately before time $y+t_i$.

Our non-parametric estimator of the aggregate integrated hazard from time y to $y+t$ is shown in equation (1). An example of $\hat{\Lambda}_{y,t}$ for the FRA portfolio is shown in Figure 2(a). The number of comparisons required for $\{l_{y+t_i^-}\}$ grows with the product of n_y and the number of lives, so we use parallel processing over 63 threads to reduce calculation time [Butenhof, 1997].

We then use the Nelson-Aalen estimator in equation (1) to create a semi-parametric estimator of the mortality hazard in time, $\hat{\mu}_{y+t}$, as per equation (2). This is a central difference around $\hat{\Lambda}_{y,t}$ with a bandwidth parameter of $c > 0$ years. The result is shown in Figure 2(b) for the FRA portfolio. Equation (2) is essentially a uniform kernel estimator, and more sophisticated kernel estimators are available; see Anderson et al. [1992, Section IV.2.1]. In this paper c is set subjectively.

$$\hat{\Lambda}_{y,t} = \sum_{t_i \leq t} \frac{d_{y+t_i}}{l_{y+t_i^-}} \quad (1)$$

$$\hat{\mu}_{y+t} = \frac{1}{c} \left(\hat{\Lambda}_{y,t+c/2} - \hat{\Lambda}_{y,t-c/2} \right) \quad (2)$$

Despite the major sources of unmodelled heterogeneity, the estimators in equations (1) and (2) are useful for revealing short-term fluctuations of mortality, as shown in Figure 2. Although $\hat{\Lambda}_{2010,t}$ looks plausibly linear, $\hat{\mu}_{2010+t}$ reveals the strong seasonality observed in many pensioner populations [Richards et al., 2020]. The overall rising trend in Figure 2(b) is due to the increasing average age of the FRA portfolio shown in Figure 1. The amplitude of seasonal fluctuations increases with the increasing average age, a phenomenon measured in this and other portfolios in Richards et al. [2020, Section 5].

Figure 2: (a) Nelson-Aalen estimate $\hat{\Lambda}_{2010,t}$, (b) estimate $\hat{\mu}_{2010+t}$ with $c = 0.5$. Mortality is revealed to be a highly time-dependent process, with material swings based on the season. Source: own calculations for FRA portfolio described in Table 1.

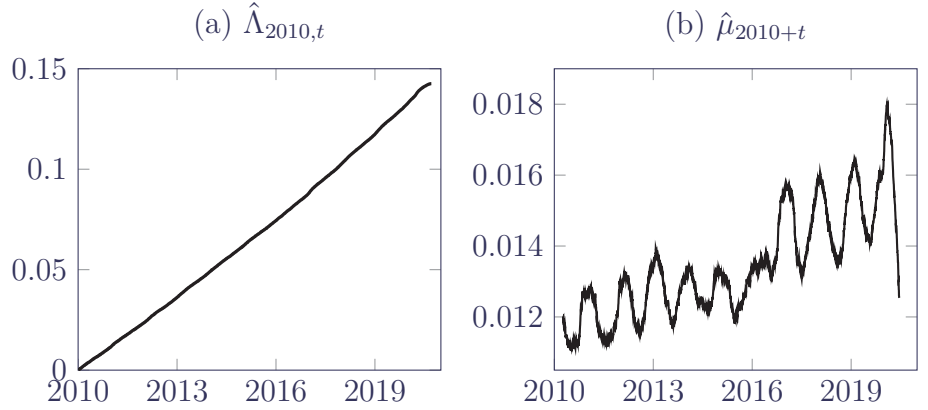
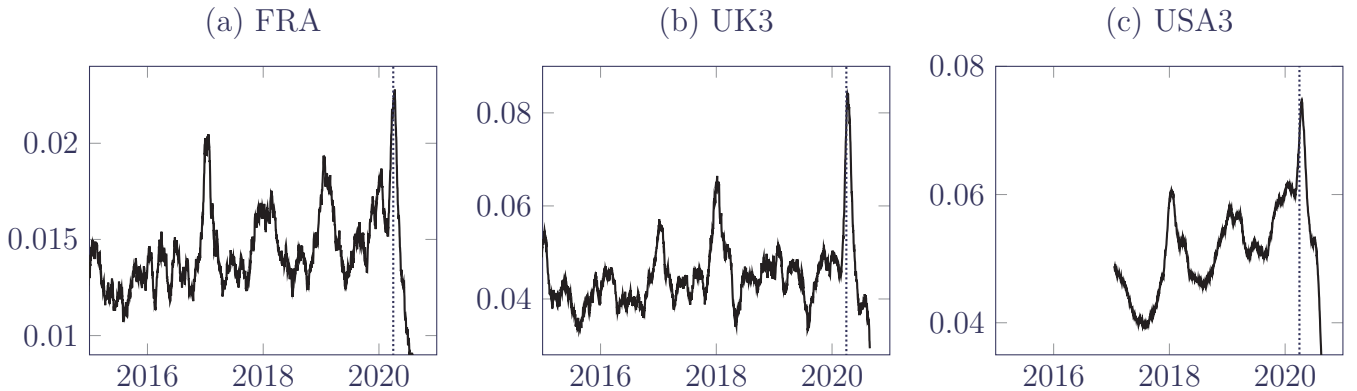


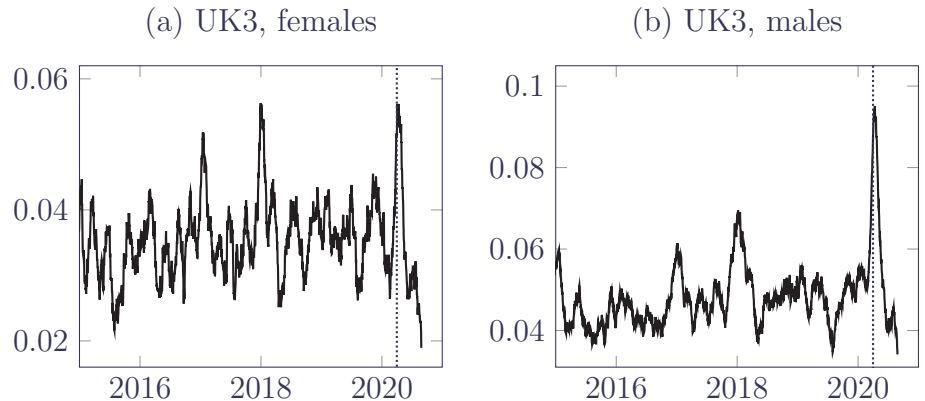
Figure 3: $\hat{\mu}_{y+t}$ for FRA, UK3 and USA3 portfolios with $c = 0.1$ and 1st April 2020 marked with a dotted vertical line. The vertical scales are different because of the younger average age of the FRA portfolio. y varies by portfolio, but only values of $y + t$ since 1st January 2015 are shown for comparability.



Equation (2) works best where there are several reported deaths per day. This applies to all three annuity portfolios here, but term-assurance portfolios would have to be larger due to the younger average age. Although the UK3 portfolio is the smallest in terms of policy counts (Table 1), its higher average age (see Figure 1) means there are more deaths per day than the much larger FRA portfolio. More deaths per day means that a portfolio can support a smaller value of c to reveal greater detail, as in Figure 3 where we can see the covid-19 mortality spike in April 2020 for the FRA, UK3 and USA3 portfolios. The spike is lower for the FRA portfolio because of the lower average age, as the mortality rate of those infected with SarsCov2 increases strongly with age; see CCAES [2020] and Istat [2020]. Figure 3 also benchmarks the covid-19 spike against past seasonal variation: for the FRA portfolio, the covid-19 spike is marginally higher than the winter period at the start of 2017, whereas for the UK3 portfolio the covid-19 spike is at least a third higher than the 2017/2018 winter. For the UK3 portfolio, the death rate peaked at 19.6 per 100,000 annuities in the second week of January 2018, whereas it peaked at 26.8 in the first week of April 2020.

Figure 4 shows $\hat{\mu}_{2015+t}$ separately for males and females in the UK3 portfolio. The females in Figure 4(a) experienced a covid-19 spike in April 2020 that is not much worse than the spike in the winter of 2017/2018. In contrast, Figure 4(b) shows that males experienced far worse covid-19 mortality, with a peak death rate of 29.8 per 100,000 in the first week of April 2020 against a prior peak death rate of 23.8 per 100,000 in the first week of January 2018. The heavier

Figure 4: $\hat{\mu}_{2015+t}$ with $c = 0.1$ separately for males and females in UK3 portfolio with 1st April 2020 marked with a dotted vertical line. The vertical scales are different because males have a higher mortality rate than females.



covid-19 mortality of males was also present in the data of CCAES [2020] and Istat [2020].

Equations (1) and (2) are useful for exploratory analysis, and graphs like Figure 3 are also useful for communication with non-experts. One practical aspect of equation (1) is that only three data items for each annuity are required for its computation: (i) the annuity commencement date, (ii) the date of cessation of observation (death, early withdrawal or the extract date), and (iii) whether the cessation is a death. This means that no personal data are required for the calculation of the mortality hazard in equation (2). This is a useful feature when modern data-protection and privacy laws restrict both the use and sharing of personally identifiable information (PII); indeed, the data for USA3 could only be shared for this research precisely because they contained no PII. Data privacy and further applications of the semi-parametric approach are discussed in Appendix A.

However, there are several drawbacks of equation (2). One is that $\hat{\mu}_{y+t}$ is not defined for the last $c/2$ years of the experience data. For smaller portfolios requiring larger values of c (see Figure 15), this reduces the usefulness of the semi-parametric estimator as a timely statement of recent mortality. Furthermore, even where $\hat{\mu}_{y+t}$ is defined, the most recently calculated values are affected by unreported deaths — the FRA and USA3 portfolios in Figure 3 show particularly pronounced drops in reported mortality levels in mid-2020 caused by such delays. We therefore need a methodology that can produce results that are both up-to-date and adjust for reporting delays. Before this, however, we first explore the nature and extent of late-reported deaths in the FRA and UK3 portfolios, as these portfolios have more than one data extract.

4 Reporting delays

Figures 3(a) and (c) show a dramatic fall in apparent mortality after the peak in April 2020, far below the level of pre-pandemic mortality. This is due to delays in the reporting and recording of deaths, i.e. occurred but not reported (OBNR) to use the terminology of Lawless [1994]. Such OBNR effects make the estimator in equation (2) less useful for the period leading up to the extract date. This is in addition to the loss of the $c/2$ years prior to the extract date due to the nature of equation (2). This gives rise to a series of practical questions: (i) how extensive is the OBNR effect, (ii) how far back does the OBNR effect matter, and (iii) can we adjust for OBNR to compensate?

We assume that a death takes a time $v > 0$ to be notified to the insurer. v is referred to as secondary data, i.e. information about the death reporting rather than about the underlying life.

For the FRA and UK3 datasets we each have two separate extracts at times u_1 and u_2 in 2020. Figure 5 shows the three different ways in which a death at time $u < u_1$ can be related to the two extracts. Type A deaths both occur and are reported before u_1 . Without separate recording of the notification date, such deaths tell us only that $v \in (0, u_1 - u]$. Type B deaths occur before u_1 but are still unreported by u_2 . Such deaths are unknown to us and so our delay data are right-truncated. Type C deaths occur before u_1 , are not reported by u_1 but are reported by u_2 ; here the delay data are interval-censored [Collett, 2003, Chapter 9] and all we can say about the reporting delay is that $v \in (u_1 - u, u_2 - u]$.

We do not know the exact date of each extract, so we assume it is the day after the last date of death. This simplifying assumption has a significant impact on the AICs in the likes of Tables 4 and 5, although it does not change the conclusion that the OBNR effect is statistically significant. Table 2 shows summary details of the late-reported deaths observed between two successive extracts of the FRA and UK3 portfolios.

Table 2 shows that the FRA portfolio has a bigger OBNR issue than the UK3 portfolio, as the ratio of Type C deaths to Type A deaths in 2020 is much higher. We also see that both portfolios have occasional delays lasting more than a decade. Figure 6 shows the ogives of the minimum reporting delays for Type C deaths, i.e. $u_1 - u$. Although Figure 6 makes it appear that the UK3 portfolio has a worse reporting delay due to the lower cumulative proportion, Table 2 shows that this is actually because the Type C deaths for UK3 are the most seriously delayed cases.

To illustrate the impact of OBNR deaths, Table 3 shows a week in June 2020 for the UK3 portfolio. As noted in Dodd et al. [2015, p108], there are numerous effects of OBNR deaths on the likes of equation (2). There are three direct ways for missing deaths to have an impact: (i) there are more distinct death times than the n_y ones observed in

Figure 5: Three deaths A, B and C occurring on the same date, u , and how their notification relates to data extracts at times u_1 and u_2 ($u < u_1 < u_2$).

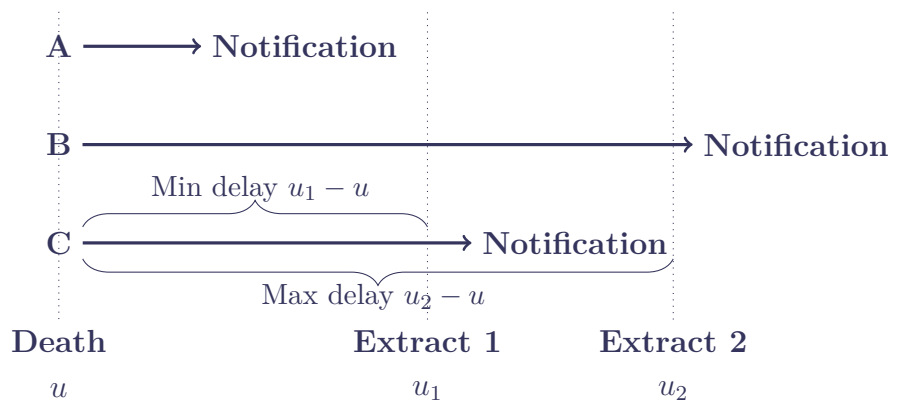
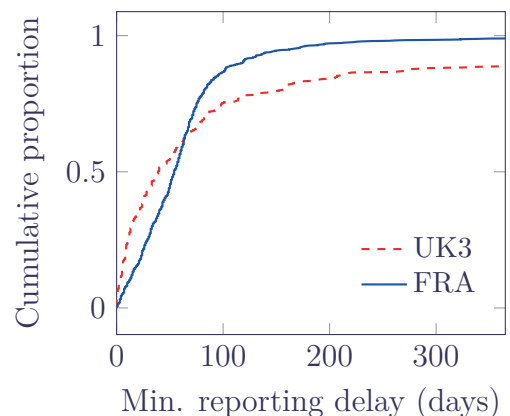


Table 2: Summary of observed reporting delays between successive extracts.

	FRA	UK3
Assumed first extract date, u_1	18 June 2020	18 June 2020
Assumed second extract date, u_2	18 Sept. 2020	16 Sept. 2020
Type A deaths in 2020 before u_1	1,171	2,045
Type C deaths notified in $[u_1, u_2)$	853	488
Ratio of Type C to Type A deaths	0.729	0.239
Type C deaths: median of $(u_1 - u)$	54 days	39 days
Type C deaths: largest of $(u_1 - u)$	15.95 years	24.48 years

Figure 6: Ogives of minimum Type C reporting delays in days. The horizontal axis is restricted, as the presence of decade-long delays otherwise distorts the plot.



a given extract, as demonstrated by the row for 16 June 2020, (ii) the d_{y+t_i} death counts are higher, and (iii) $l_{y+t_i}^-$ counts are lower due to other late-reported deaths for the period before 11 June 2020. This effect stretches back months prior to the extract, as can be seen in the in-force policy counts at 1 April 2020 in Table 1. Points (ii) and (iii) both act on equation (2) to progressively under-state the mortality hazard closer to the extract date.

For portfolios like UK3 there is an indirect fourth impact on policy counts from late-reported deaths: upon the death of a primary annuitant, a surviving spouse’s annuity may be set up. In addition to past death counts increasing due to reporting delays, policy counts may also increase if a back-dated spouse’s annuity is set up. Such a case would lead to d_{y+t_i} increasing by 1 at the time of death of the primary annuitant, but the corresponding $l_{y+t_i}^-$ count would be un-

Table 3: Impact of late-reported deaths from comparing two extracts of UK3 portfolio in Table 1. The later extract not only contains more deaths due to reporting delays, but reduced in-force counts as well. In the June 2020 extract there were no deaths on 16 June and so no policy count was calculated for that date.

Date	June 2020 extract:			Sept. 2020 extract:		
	In-force policies	Deaths	New policies	In-force policies	Deaths	New policies
2020-06-11	145,166	6	3	144,934	18	4
2020-06-12	145,163	3	8	144,920	16	14
2020-06-13	145,168	9	0	144,918	14	5
2020-06-14	145,159	1	4	144,909	7	4
2020-06-15	145,162	3	1	144,906	15	7
2020-06-16	n/a	n/a	8	144,898	8	12
2020-06-17	145,168	3	1	144,902	29	8

changed due to the new spouse’s annuity offsetting the main annuitant’s death. The columns of new policy counts in Table 3 illustrates the impact of retrospective spouse annuities being set up.

The impact of OBNR can be expressed as a function, $R(s, u_1, u_2)$, of the time $s > 0$ prior to the extract at u_1 . $R(s, u_1, u_2)$ is the ratio

$$R(s, u_1, u_2) = \frac{\hat{\mu}_{u_1-s} \text{ using extract at time } u_1}{\hat{\mu}_{u_1-s} \text{ using extract at time } u_2} \quad (3)$$

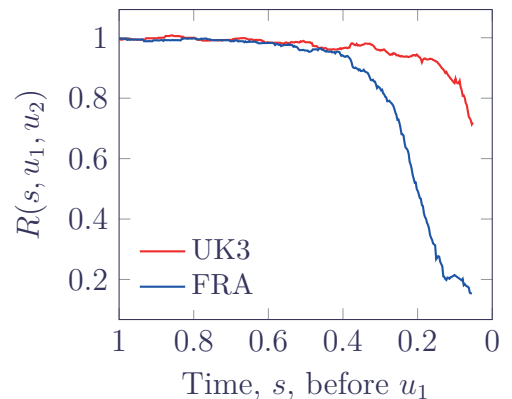
of the $\hat{\mu}_y$ estimates using extracts at times u_1 and u_2 , as in equation (3). R is only defined for $s > c/2$ and is plotted in Figure 7 for the FRA and UK3 portfolios. Both portfolios happen to have extracts for June 2020 (u_1) and September 2020 (u_2) and there is a steady fall in reported mortality in the months leading up to the June 2020 extract in each case (dramatically so for the FRA portfolio).

Figure 7 shows that OBNR most affects the recent mortality estimates. Indeed, the most recent FRA ratio in Figure 7(a) is probably unreliable due to very small numbers of deaths. This is partly due to different insurer administration processes, but also to national differences in death registration and probate handling. In the UK, for example, insurers can use monthly feeds of deaths from the General Registrar’s Office to match possible deaths among their annuitants instead of waiting for notification.

The semi-parametric estimates of OBNR in Figure 7 are not wholly smooth and they are undefined for the period $c/2$ years prior to the extract at time u_1 . R is also dependent on the second extract at time u_2 picking up all

the late-reported deaths, which is at best an approximation due to Type B deaths — Table 2 shows that decade-long delays can occur and Figure 6 shows that material delays exist beyond a year for

Figure 7: $R(s, u_1, u_2)$ for FRA and UK3 portfolios. The horizontal axis is reversed.



the UK3 portfolio. A better approach would be a parametric model, especially one that operated on a single data extract.

5 A parametric approach

We first note that we are interested in reporting delays only (OBNR), rather the combined delay-and-cost insurance models (IBNR) of Haastrup and Arjas [1996] and Dodd et al. [2015]. Such IBNR models are “cost-orientated, discrete-time models” [Jewell, 1989], whereas we require a pure delay model operating in continuous time.

There are numerous delay-only models, including for AIDS surveillance [Lawless, 1994], HIV incubation time [Kalbfleisch and Lawless, 1991], cancer diagnosis and reporting to a central registry [Midthune et al., 2005] and even delays in detecting bugs in software [Jewell, 1985]. A common theme in these models is the availability of secondary data, specifically the date of notification of death in addition to the date of death itself. Delay data are necessarily right-truncated [Midthune et al., 2005] and in some cases missing dates of notification can be accurately determined retrospectively [Kalbfleisch and Lawless, 1991]. However, in the examples of the FRA and UK3 portfolios our delay data are additionally interval-censored [Collett, 2003, Chapter 9] as well as right-truncated.

Jewell [1989] presents four classifications based on the availability of secondary data on delays and concludes that categories with missing dates of notification are “uninformative and useless” with respect to Bayesian methods. However, interval-censored delays are not entirely missing, and are therefore neither uninformative nor useless, as shown in the distribution of minimum delays in Figure 6. Nevertheless, inference using interval-censored, right-truncated data is tricky, and for many insurers taking two extracts might not be practical. We therefore seek an approach that can infer OBNR from a single data extract.

We model OBNR-affected observed mortality, $\mu_{x,y}^{OBNR}(u_j)$, which is related to actual mortality, $\mu_{x,y}^*$, as in equation (4). $\mu_{x,y}^{OBNR}(u_j) = \mu_{x,y}^* \rho(u_j - y, \lambda_1)$ (4) $\rho(s, \lambda_1)$ is an OBNR scaling factor reflecting the fact that not all mortality is reported in a timely manner. s is the positive-valued time before the date of extract, u_j , and λ_1 is the decay parameter reflecting how quickly the OBNR effect changes with s . $\rho(s, \lambda_1)$ implicitly allows for unreported deaths of both Type B and Type C in Figure 5 by modelling the shortfall in mortality as y approaches the extract date, u_j (see for example the shortfall demonstrated by the large negative residuals in Figure 10(a)). In most ordinary business circumstances there will just be a single extract ($j = 1$), but we use the index j to allow for the circumstances where a second extract is available, as for the FRA and UK3 portfolios. We note an important difference between equation (4) on one hand and existing IBNR models and pure delay models on the other: the OBNR in our data is a nuisance effect and we only want to estimate it to eliminate it. Both the function ρ and the parameter λ_1 are therefore used in estimation, but neither plays any role in applications of the parameterised model $\mu_{x,y}^*$, such as pricing or reserving.

Some candidate functions for ρ are listed in equations (5)-(10) and plotted in Figure 8 for a specimen value of λ_1 . Figure 8(a) shows the variety of shapes possible, allowing the choice of function and parameter λ_1 to fit the delay shape of a given portfolio. Figure 8(b) shows the role of the decay parameter, λ_1 , with lower values meaning more OBNR.

$$\text{Logistic} : se^{\lambda_1} / (1 + se^{\lambda_1}) \quad (5)$$

$$\text{Exponential} : 1 - \exp(-se^{\lambda_1}) \quad (6)$$

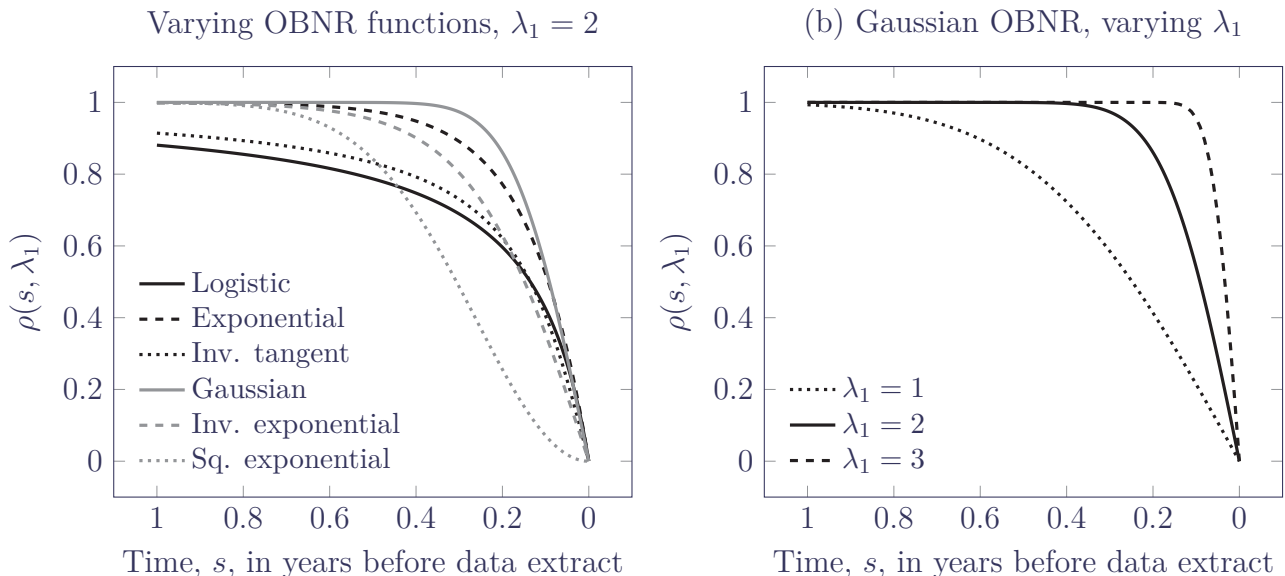
$$\text{Inverse tangent} : \frac{2}{\pi} \tan^{-1}(se^{\lambda_1}) \quad (7)$$

$$\text{Gaussian} : 2\Phi(se^{\lambda_1}) - 1 \quad (8)$$

$$\text{Inverse exponential} : 2 \left(1 + \exp(-se^{\lambda_1}) \right)^{-1} - 1 \quad (9)$$

$$\text{Squared exponential} : 1 - \exp(-s^2 e^{\lambda_1}) \quad (10)$$

Figure 8: (a) OBNR functions with $\lambda_1 = 2$, (b) Gaussian OBNR function with varying λ_1 . The horizontal axis is reversed for comparison with Figure 7.



One aspect of equation (4) is vulnerability to confounding [Midthune et al., 2005, p69]. The choice of functions that are close to 1 until near the end of the exposure period forces the parameter λ_1 to pick up mostly OBNR effects. However, confounding of OBNR effects with seasonal effects could happen when $s < 0.5$, say because u_j was in the summer and so mortality would be falling from the winter peak [Richards et al., 2020]; equally, if u_j were in the winter, the rising mortality since the summer would lead to under-stated OBNR effects. For $s \gg 0.5$ there is also the risk of confounding OBNR effects with a time trend [Midthune et al., 2005, p62]. To reduce the impact of confounding, any model attempting to allow for OBNR effects should include both a time trend and a seasonal component, albeit strong correlations may remain between estimates (see Table 9 in Appendix B). To address the seasonal aspect of mortality, we use the cosine model from Richards et al. [2020, Section 8], as shown in equation (11):

$$\log \mu_{x,y}^* = \log \mu_{x,y} + e^\zeta \cos(2\pi(y - \tau)) \quad (11)$$

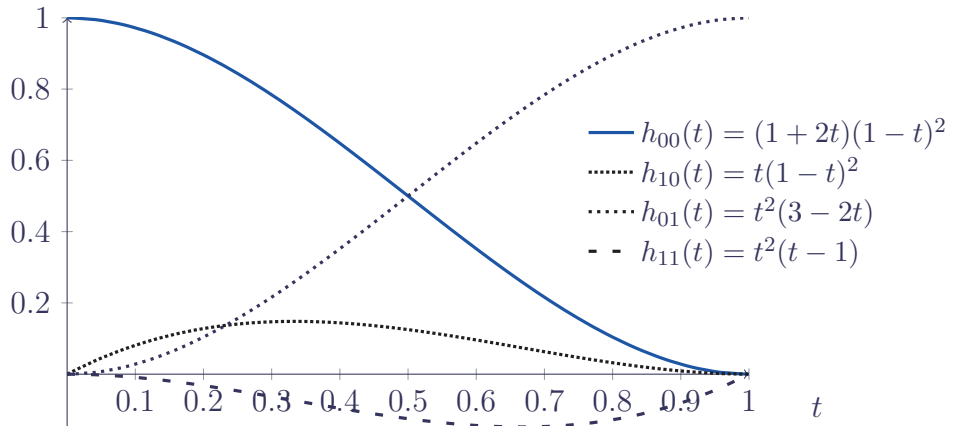
$\tau \in [0, 1)$ represents the proportion of the year after January 1st when mortality peaks and e^ζ is the peak additional mortality at that time (on a logarithmic scale). The two-parameter model for seasonal mortality in equation (11) simultaneously identifies (i) the amplitude of the average-to-peak seasonal variation (e^ζ) and (ii) the point after January 1st corresponding to the winter mortality peak (τ). The full trough-to-peak variation is $2e^\zeta$, and the definition of equation (11) forces τ to identify the winter peak, as opposed to the summer trough; we also assume that the peak is coincident in each of the years covered by a particular data set. $\mu_{x,y}$ is the non-seasonal component to the mortality hazard, i.e. varying in age and time only. The specific form of $\mu_{x,y}$ will depend on the age range under study, $x = x_0, \dots, x_1$ say. Here we will use a variant of the Hermite II model of Richards [2020] shown in equation (12):

$$\log \mu_{x,y} = (\alpha + \delta(y - 2000))h_{00}(t) + m_0h_{10}(t) + \omega h_{01}(t) \quad (12)$$

where $t = (x - x_0)/(x_1 - x_0)$ and the h functions are the cubic Hermite polynomials [Kreyszig, 1999, p868] shown in Figure 9.

To fit a survival model to individual data we maximise the log-likelihood function in equation (13) [Macdonald et al., 2018, Section 5.3] for the OBNR-affected mortality hazard, $\mu_{x,y}^{OBNR}(u_j)$, at age x at calendar time y using an extract at time u_j . Each life i enters observation at age x_i at calendar time y_i and is observed for t_i years. d_i is an indicator variable taking the value 1 if life i is reported to have

Figure 9: Hermite basis splines for $t \in [0, 1]$.



died at age $x_i + t_i$, or 0 otherwise. The potential for bias due to OBNR lies in the fact that cases with $d_i = 1$ are by definition unaffected by reporting delays, while cases with $d_i = 0$ may be. $H_{x,y}(t; u_j)$ is the integrated hazard function in equation (14). We refer to Richards et al. [2020, Appendix A] for details of numerical integration of seasonally fluctuating hazard functions.

Equation (13) is the log-likelihood for a survival model with left-truncated data; see Macdonald et al. [2018, Section 4.3]. This contrasts with survival models used in medical research, where left-truncation is rare and where likelihoods are usually for non-left-truncated data [Collett, 2003, Chapter 6]. Note also that the structure of equation (4) means that OBNR effects are allowed for without having to deal with the interval-sensored, right-truncated nature of reported delays.

$$\ell = - \sum_{i=1}^n H_{x_i, y_i}(t_i; u_j) + \sum_{i=1}^n d_i \log \mu_{x_i, y_i}^{OBNR}(u_j) \quad (13)$$

$$H_{x,y}(t; u_j) = \int_0^t \mu_{x+s, y+s}^{OBNR}(u_j) ds \quad (14)$$

6 Effectiveness of parametric OBNR model

In this section we look at the effectiveness of the parametric OBNR model in terms of model fit. Figure 10 shows the deviance residuals [McCullagh and Nelder, 1989, p39] without an OBNR term (upper panels, $\rho(s) = 1$) and with a Gaussian OBNR term (lower panels). Although both portfolios have large negative residuals for the experience closest to June 2020 due to OBNR, Figure 10(a) shows that the FRA portfolio has a more severe OBNR effect with very large negative residuals. Figure 10(d) also shows that the UK3 portfolio has been affected most strongly by covid-19 mortality, with a residual of +10 for the fourth month of 2020 (+10.3 without the OBNR term and +9.8 with it). The three next largest positive residuals in Figure 10(d) are all for January, reflecting (i) the excess winter mortality beyond the simple cosine model of equation (11) and (ii) that the covid-19 mortality is worse than the worst recent winter excess. The lower value of $\hat{\lambda}_1$ for FRA indicates a slower OBNR decay rate, and thus a larger OBNR effect.

Figure 10: Monthly deviance residuals from model fits to June 2020 data for FRA and UK3 portfolio with and without a Gaussian OBNR term. Note the change in vertical scale from (a) to (c) and (b) to (d) as the large negative residuals are eliminated.

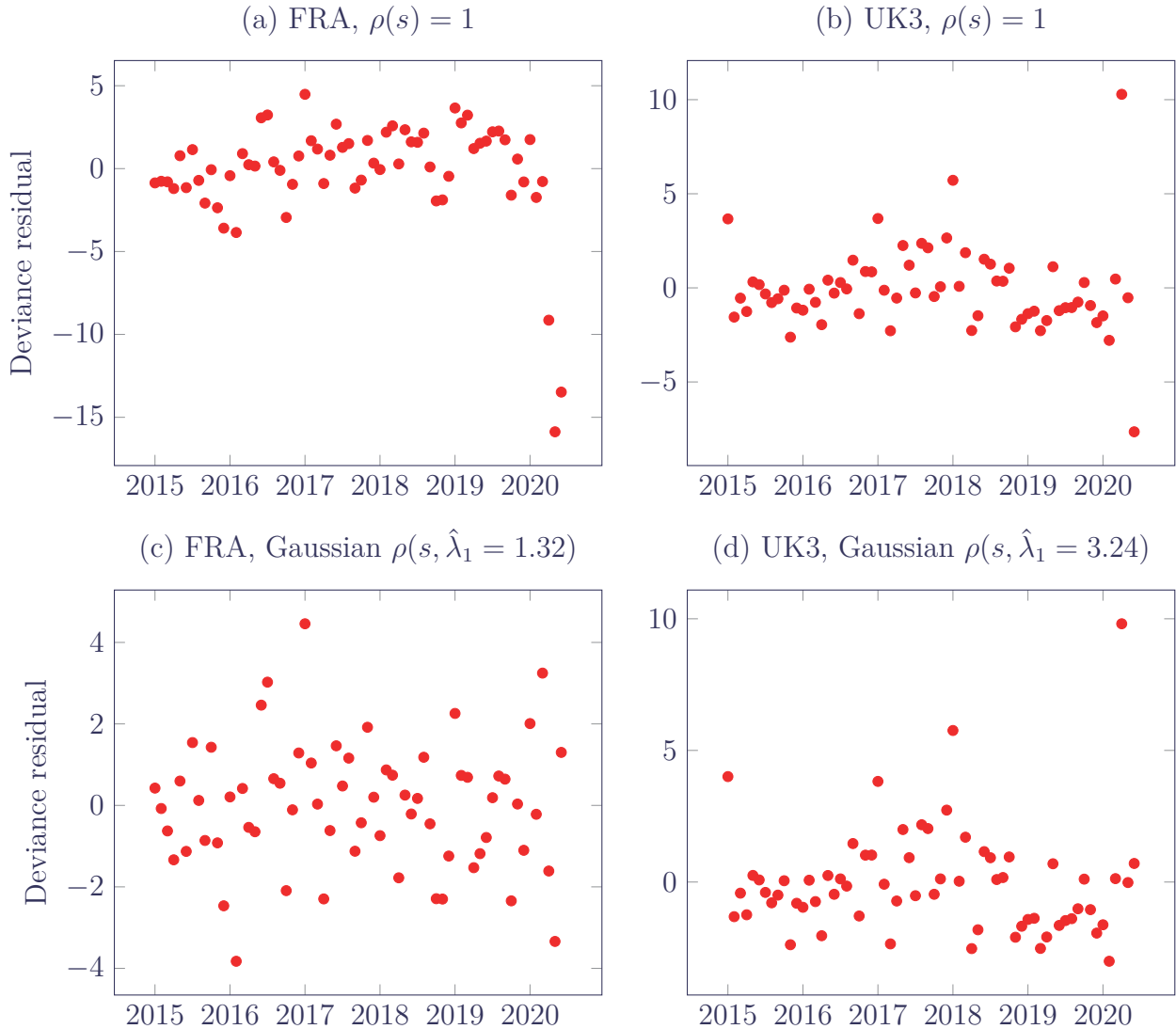


Table 4 shows the stepwise development of the AIC [Akaike, 1987], dispensing with the small-sample correction as the number of lives is very large [Macdonald et al., 2018, page 98]. For both portfolios the inclusion of a time trend makes the smallest improvement in model fit, while the larger change in fit due to the OBNR effect in the FRA portfolio reflects the stark change between panels (a) and (c) in Figure 10. The OBNR functions in equations (5)-(10) do not allow for a (quite reasonable) link between reporting delays and age. However, they do allow for interactions between λ_1 and categorical factors, and we find that neither the FRA nor UK3 portfolios has significant variation in OBNR by gender.

Table 4: Development of AIC for models fitted stepwise to experience data from 1st Jan. 2015 to June 2020.

Model factors	FRA	UK3
Age only	165,588	166,769
+Gender	-736	-518
+Time	-79	-34
+Season	-172	-156
+OBNR (Gaussian)	-593	-93
Full model	164,008	165,968

Table 5 shows the differences in fit for the various OBNR models in equations (5)–(10). Although these two portfolios differ in the extent and timing of their OBNR, the various OBNR functions have a similar order with respect to the AIC, with the Gaussian OBNR model providing the best fit (or best-equal fit). The exception is the squared-exponential OBNR model, which failed to fit properly as the seasonal component of the model wouldn't fit. Refitting the model with $m_0 = 0$ in equation (12) however worked, suggesting an occasionally problematic interplay between parameters.

Table 5: AICs for OBNR models fitted to experience data from 1st January 2015 to June 2020.

OBNR model	FRA	UK3
Gaussian	164,008	165,968
Inverse exponential	164,010	165,968
Exponential	164,025	165,971
Squared exponential	failed to fit	165,976
Inverse tangent	164,033	165,992
Logistic	164,062	166,001

7 Parametric OBNR models as a forecast of delays

In this section we look at the ability of a parametric OBNR model fitted to a single data extract to predict the pending delays. For this we will fit the model for the OBNR-affected mortality of Section 5 to the data for the June extract, then we compare these OBNR functions to the effect of pre-June deaths reported by September. For the actual delay effect we will use the semi-parametric R function of equation (7); this will not be a complete statement of OBNR due to still-unreported Type B deaths, but it will be a useful comparison.

Figure 11(a) shows the fitted OBNR curves, $\rho(s, \lambda_1)$, for the FRA portfolio, while Figure 11(b) shows the best-fitting Gaussian OBNR curve versus the semi-parametric $R(s, u_1, u_2)$ from Figure 7(a). Figure 11(b) shows that it is possible to make workable estimates of mortality OBNR for a portfolio based on just a single extract of data, i.e. without any secondary data on the delays themselves — the fitted OBNR adjustment based on data up to time u_1 follows the broad shape and trajectory of the OBNR in the subsequent experience after u_1 . In this sense the OBNR function is actually a short-term forecast based on patterns leading up to the extract date.

In contrast, Figure 12(b) shows that for the UK3 portfolio the best-fitting of the OBNR functions in June proved to be a poorer forecast of the actual experience represented by the semi-parametric estimate using the late-reported deaths to September. In fact, the two poorest-fitting OBNR functions would have done a better job. This contrast can perhaps be explained by the fact that the UK3 portfolio has a much smaller OBNR issue — Table 2 shows that the UK3 portfolio has a much lower proportion of Type C deaths compared to the FRA portfolio. A possible reason for the better performance of the parametric OBNR function for the FRA portfolio is that the OBNR issue is both relatively larger and affects a longer period of time leading up to the extract. This is suggested by the pattern of residuals in Figure 11(a). The parametric OBNR functions therefore work best where they are most needed. Equally, where they are least needed, the precise choice of OBNR function is less critical.

Figure 11: (a) Estimated OBNR functions for June 2020 extract for FRA portfolio, (b) comparison of June OBNR estimates with OBNR function with lowest AIC against subsequent estimated OBNR using September 2020 extract. Panel (b) shows that the best-fitting parametric OBNR function using the June extract proved to be a workable overall forecast of the OBNR; this is despite the fact that the precise trajectory of the actual OBNR lies outside the 95% confidence envelope for the estimated OBNR function. The horizontal axes are reversed.

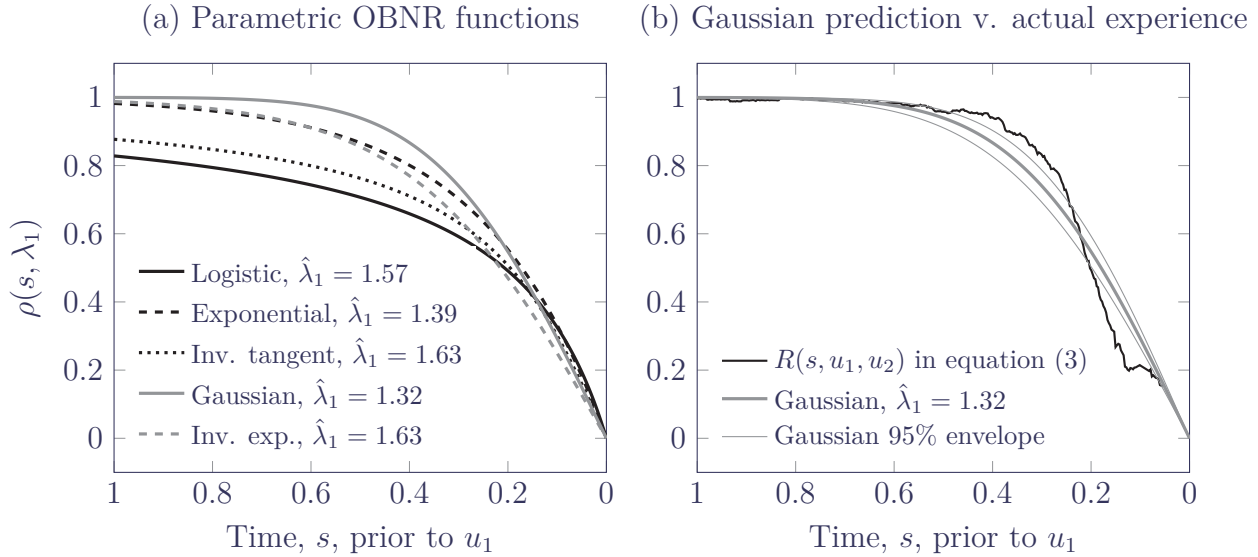
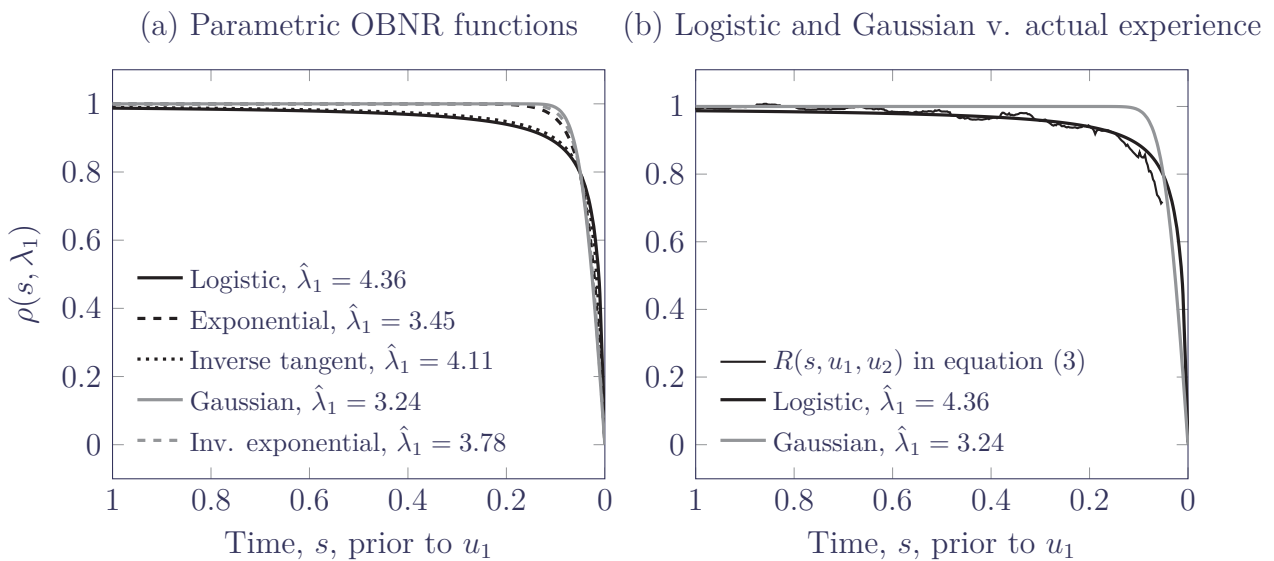
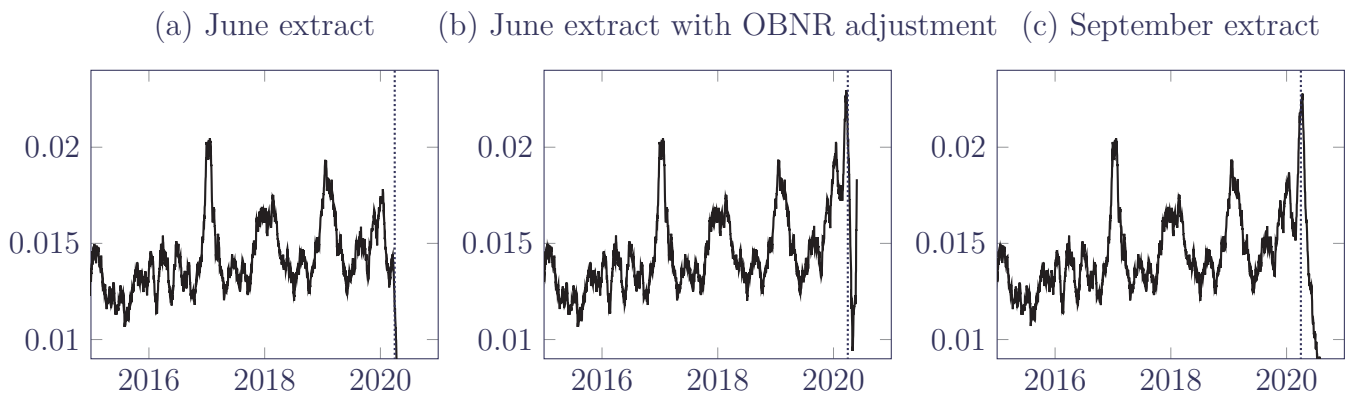


Figure 12: (a) Estimated OBNR functions for June 2020 extract for UK3 portfolio, (b) comparison of OBNR estimates with highest and lowest AICs against subsequent estimated OBNR using September 2020 extract. The OBNR functions with the joint-equal lowest AIC in the June extract proved to be poor projections of what the estimated OBNR was using the September extract; in fact the OBNR functions with the highest AIC (logistic and inverse tangent) would have been better forecasts. The horizontal axes are reversed.



The final step is to make use of the fitted OBNR function to estimate current mortality. An illustration of this is given in Figure 13: panel (a) shows $\hat{\mu}_{2015+t}$ from equation (2) using the extract in June 2020, where the OBNR issue means that the covid-19 shock in April 2020 (marked with the vertical dotted line) is entirely absent. Panel (b) shows $\hat{\mu}_{2015+t}/\rho(s, \hat{\lambda}_1)$ for the Gaussian ρ function in equation (8), where $\hat{\lambda}_1 = 1.32332$ (s.e. 0.0508) is estimated from the same June 2020 extract; the resulting adjustment makes little difference to the pre-2020 mortality, but both the January winter peak and the covid-19 spike now appear. Panel (c) shows $\hat{\mu}_{2015+t}$ calculated using the extract from September 2020, i.e. with three months additional reporting of delayed deaths. We can see that the OBNR adjustment in Figure 13(b) has done a creditable job of predicting the covid-19 spike in April 2020, albeit it has slightly over-stated the January 2020 winter peak. The parametric OBNR model therefore allows actuaries to estimate OBNR effects from a single extract, thus providing up-to-date short-term forecasts of what emerging portfolio mortality levels actually are, even in the face of material reporting delays.

Figure 13: $\hat{\mu}_{2015+t}$ for FRA portfolio with $c = 0.1$: (a) calculated using June 2020 extract; (b) as (a) but adjusted using Gaussian OBNR function estimated from June 2020 extract; (c) calculated using September 2020 extract.



8 Conclusions

Portfolio data handled by actuaries contain individual records observed over a number of years. This paper presents a semi-parametric estimator for the mortality hazard that illustrates short-term fluctuations in time, such as seasonal variation and the covid-19 mortality shock. Using this estimator we find that the initial covid-19 shock peaked in the same week in April 2020 for three separate portfolios in France, the UK and the US. The visual nature of the estimator makes it a practical tool not just for actuarial work analysing portfolio data, but also in communicating with non-specialists.

However, the semi-parametric estimator has various limitations, not least its vulnerability to reporting delays for deaths. We present a parametric model that allows for the effect of these delays and quantify the extent and impact of such delays on the most recent stated mortality rates. The model permits the estimation of an OBNR adjustment from a single data extract, thus providing a means for continuous, up-to-date reporting. When combined with a model allowing for seasonal variation, this allows actuaries to quantify mortality effects as they happen, while providing short-term forecasts for planning and taking management action.

Acknowledgments

The author thanks the following for their support in providing data: Alex Oleksandr Zavershynskyy; Clément Frappier and H el ene Queau; Joseph England and Anil Gandhi; Denis Dupont and Murray Wright. The author also thanks Gavin Ritchie, Torsten Kleinow, Stefan Ramonat and Angus Macdonald for helpful comments on earlier drafts of this paper. Any errors or omissions remain the responsibility of the author. Calculations were performed using the Longevitas survival-modelling system and bespoke programs written in R and C++. Graphs were produced in tikz and pgfplots, while typesetting was done in L ATEX.

References

- H. Akaike. Factor analysis and AIC. *Psychometrika*, 52:317–333, 1987. ISSN 0033–3123. doi: <https://doi.org/10.1007/BF02294359>.
- P. K. Anderson,  . Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer, New York, 1992. ISBN 0-387-94519-9.
- D. R. Butenhof. *Programming with POSIX Threads*. Addison-Wesley, Boston, 1997. ISBN 978-0-201-63392-4.
- CCAES. Actualizaci on no. 120. Enfermedad por el coronavirus (COVID-19). 29.05.2020. Technical Report 120, Centro de Coordinaci on de Alertas y Emergencias Sanitarias, May 2020.
- D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall, Boca Raton, second edition, 2003. ISBN 1-58488-325-1.
- Council of European Union. Regulation (EU) 2016/679. *Official Journal of the European Union*, 59, 2016. ISSN 1977-0677.
- V. A. B. Djeundje and I. D. Currie. Smoothing dispersed counts with applications to mortality data. *Annals of Actuarial Science*, 5(I):33–52, 2011.
- E. Dodd, G. Streftaris, H. R. Waters, and A. D. Stott. The effect of model uncertainty on the pricing of critical illness insurance. *Annals of Actuarial Science*, 9(1):108–133, 2015. doi: 10.1017/S1748499514000244.
- S. Haastруп and E. Arjas. Claims reserving in continuous time; a nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2):139–164, 1996. doi: 10.2143/AST.26.2.563216.
- Istat. Total deaths per age class, week of demographic event and municipality of residence at the time of death. Technical report, Istituto Nazionale di Statistica, 2020. URL <https://www.istat.it/en/archivio/240106>.
- W. S. Jewell. Bayesian extensions to a basic model of software reliability. *IEEE Transactions on Software Engineering*, SE-11(12):1465–1471, 1985. doi: 10.1109/TSE.1985.231890.
- W. S. Jewell. Predicting IBNYR Events and Delays: I. Continuous Time. *ASTIN Bulletin*, 19(1): 25–55, 1989. doi: 10.2143/AST.19.1.2014914.
- J. D. Kalbfleisch and J. F. Lawless. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statistica Sinica*, 1(1):19–32, 1991. ISSN 10170405, 19968507. URL <http://www.jstor.org/stable/24303991>.

- E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.
- E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley and Sons, eighth edition, 1999. ISBN 0-471-33328-X.
- J. F. Lawless. Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics*, 22(1):15–31, 1994. doi: <https://doi.org/10.2307/3315826.n1>.
- A. S. Macdonald, S. J. Richards, and I. D. Currie. *Modelling Mortality with Actuarial Applications*. Cambridge University Press, Cambridge, 2018. ISBN 978-1-107-04541-5.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, second edition, 1989. ISBN 0-412-31760-5.
- D. N. Midthune, M. P. Fay, L. X. Clegg, and E. J. Feuer. Modeling reporting delays and reporting corrections in cancer registry data. *Journal of the American Statistical Association*, 100(469): 61–70, 2005. ISSN 01621459. URL <http://www.jstor.org/stable/27590519>.
- Minister of Justice for Canada. Personal Information Protection and Electronic Documents Act. S.C. 2000, c. 5, 2018.
- W. Nelson. Theory and applications of hazard plotting for censored failure times. *Technometrics*, 14:945–965, 1958.
- R. Rau. *Seasonality in human mortality: a demographic approach*. Demographic Research Monographs 03 XVI. Springer Verlag, 2007. ISBN 3-540-44900-0.
- S. J. Richards. A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal*, 2020:2:110–127, 2020. doi: 10.1080/03461238.2019.1642239.
- S. J. Richards, K. Kaufhold, and S. Rosenbusch. Creating portfolio-specific mortality tables: a case study. *European Actuarial Journal*, 3 (2):295–319, 2013. doi: <https://doi.org/10.1007/s13385-013-0076-6>.
- S. J. Richards, S. J. Ramonat, G. Vesper, and T. Kleinow. Modelling seasonal mortality with individual data. *Scandinavian Actuarial Journal*, pages 1–15, 2020. doi: 10.1080/03461238.2020.1777194.
- State of California Office of Legislative Counsel. The California Consumer Privacy Act of 2018. Assembly Bill No. 375, 2018.
- The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. *China CDC Weekly*, 2:113, 2020. ISSN 2096-7071. URL <http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51>.

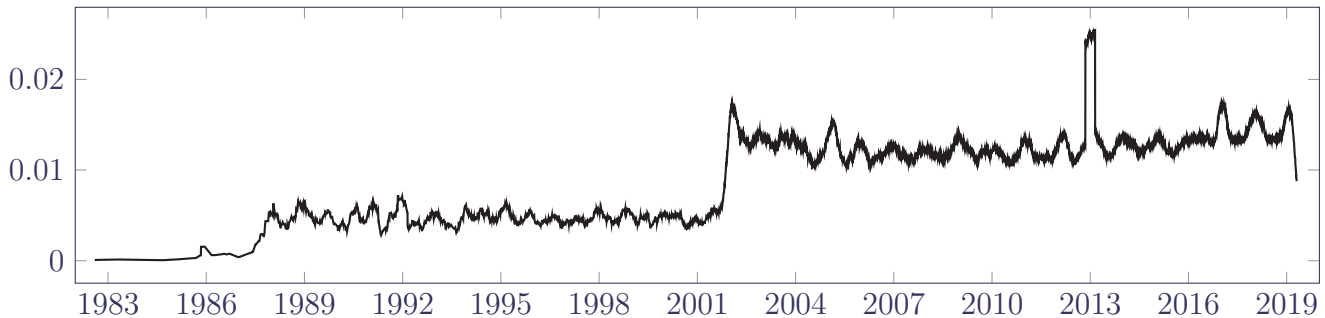
Appendices

A Detecting data issues with the semi-parametric estimator

The estimator in equation (2) is primarily for investigating short-term fluctuations in time, especially mortality spikes and seasonal variation. However, we have found that the estimator also has application to the identification of data-quality issues. This appendix demonstrates some examples.

Figure 14 shows that the June 2019 extract for the FRA portfolio in Table 1 has three phases: (i) some sparse deaths data for a period up to around 1987 that clearly cannot be used for mortality analysis, (ii) a period of stable experience over 1988–2001, and (iii) a period post-2002 also with stable experience, but with a far higher aggregate mortality rate. The first question is what caused the jump between the second and third phases? We could be dealing with a portfolio that was merged with another portfolio with a substantially higher age, or else the removal of a block of annuities with higher average age as in Figure 1 for UK3. However, in this case the policy counts represented by $l_{y+t_i^-}$ (not shown) demonstrate steady growth and no discontinuity, which means that the 1988–2001 phase of experience data must be systematically missing death records, regardless of its apparent stability. The estimator in equation (2) reveals the usable part of the exposure period to be post-2002 for the FRA portfolio.

Figure 14: Plot of $\hat{\mu}_{1982+t}$ using June 2019 extract for FRA portfolio in Table 1. Source: own calculations using equation (2) with $c = 0.3$. Only the period from 2002 onwards has complete death data for analysis, although the spike at the start of 2013 suggests some invalidly processed deaths.



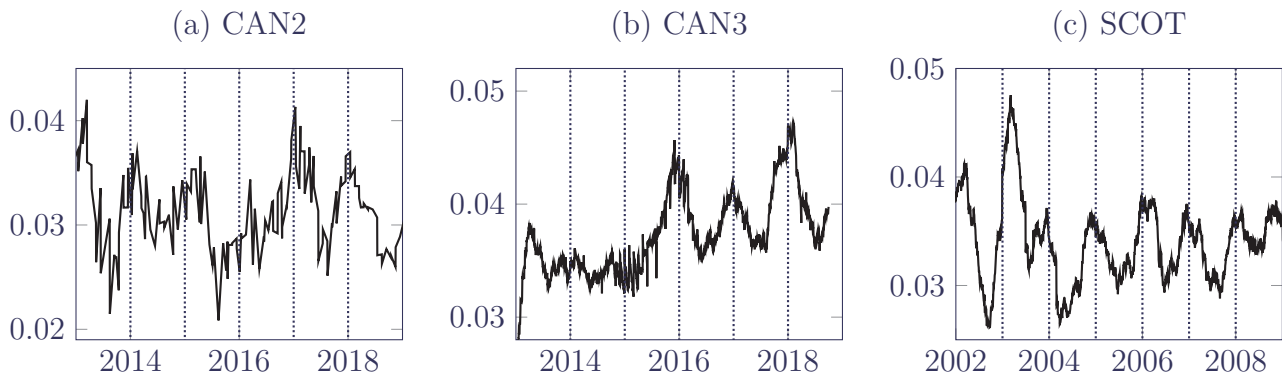
The spike in mortality rate at the start of 2013 in Figure 14 is also of interest. This was traced to 670 “deaths” on 1st January 2013, which was clearly not the actual date of death for these annuitants. This sort of data issue would also be caught by the tabulation of the most commonly occurring field values [Macdonald et al., 2018, Section 2.7]. However, the estimator in equation (2) provides a clear visual indicator that something is amiss with the data at the start of 2013.

The estimator in equation (2) can also highlight periods of potentially suspect data beyond the two examples in Figure 14. Table 6 summarises the data available for three pension schemes of roughly comparable size, while Figure 15 shows the estimator in equation (2) for each. Figure 15(a) shows what happens when the dates of death are not accurately recorded — the unclear signal is due to most deaths for CAN2 being recorded as happening in the middle of each month, rather than on the actual date of death. Figure 15(b) shows that the expected seasonal pattern is present for CAN3 in 2016–2019, but that it is missing from late 2013 to early 2015; the record counts are consistent and have no discontinuity (not shown), so something is probably amiss with the death recording prior to mid-2015. In contrast, Figure 15(c) shows a reliable seasonal pattern for the SCOT portfolio, indicating that the data have no obvious time-based issues.

Table 6: Overview of three medium-sized pension schemes, none of which contains mortality experience of the covid-19 pandemic. The naming convention was chosen for consistency with Richards et al. [2020], which uses two of the same portfolios.

Country	Portfolio	Description	Max. lives	Total deaths
Canada	CAN2	Single-employer defined-benefit occupational pension scheme.	9,967	2,953
	CAN3	Industry-wide defined-benefit occupational pension scheme for blue-collar workers.	26,494	5,989
Scotland	SCOT	Defined-benefit occupational pension scheme for a single Scottish local authority.	13,638	3,732

Figure 15: Plot of $\hat{\mu}_{2013+t}$ for CAN2 and CAN3 portfolios and $\hat{\mu}_{2002+t}$ for SCOT in Table 6. Source: own calculations using equation (2) with $c = 0.5$. The lack of a clear signal for CAN2 is caused by most deaths being logged as occurring in the middle of the month. The CAN3 portfolio is missing the expected seasonal pattern from late 2013 to early 2015, raising questions over the accuracy of the experience data during that period. In contrast, the SCOT portfolio has a clear recurring seasonal signal throughout, indicating that the data have no obvious time-based issues for the period shown.



A key step in reinsurance transactions is the validation of the data available, and individual policy records enable particularly thorough data-quality checks [Macdonald et al., 2018, Chapter 2]. However, data-protection and privacy laws in many territories often restrict the sharing and processing of personal data; examples include GDPR in the EU and UK [Council of European Union, 2016], CCPA in California [State of California Office of Legislative Counsel, 2018] and PIPEDA in Canada [Minister of Justice for Canada, 2018]. Data-protection laws typically apply to personally identifiable data, i.e. data that could be used to identify a living person (Canadian laws are even stricter and further protect the personal data of the recently deceased). Such laws sharply reduce the detail in data shared between counterparties for bulk annuities, longevity swaps and reinsurance treaties. This unfortunately also limits those counterparties' ability to validate the experience data. Particularly problematic are cases where risk, and therefore data, are transferred between jurisdictions, e.g. from an insurer covered by the EU's GDPR directive to a reinsurer outside the European Union. This forces the consideration of new techniques and tools for checking the validity of mortality-experience data. As discussed in Section 2, the estimator in equation (2) requires no personal data for its computation. There is therefore no reason for a risk cedant not to pass over the minimum data fields required, making equation (2) a useful data-quality check in an environment of restricted data sharing.

B Parameters

There are two kinds of parameters to be set for a Hermite model [Richards, 2020] of mortality: (i) configuration parameters, whose values are decided in advance by the analyst, and (ii) parameters whose values are estimated from the data.

B.1 Parameters set by the analyst

Table 7 sets out the configuration parameters that need to be set in advance by the analyst, i.e. they are not estimated from the data. The values used in the main body of the paper are given.

Table 7: Configuration parameters for the Hermite model family.

Parameter	Value	Description and role
x_0	50	Age below which $\log \mu_x$ is deemed constant in age; see equation (12).
x_1	110	Age above which $\log \mu_x$ is deemed constant in age; see equation (12).

In addition, the analyst must also choose an OBNR function from equations (5)–(10).

B.2 Parameters estimated from the data

Table 8 sets out the parameters whose values are estimated from the data by maximum likelihood.

Table 8: Overview of seasonal and OBNR parameters.

Parameter	Name	Description and role of parameter
α	Intercept	$\log \mu_{x_0}$, the minimum mortality hazard at age x_0 and below.
m_0	AgeGradientYoungest	Initial rate of change of $\log(\text{mortality})$ with increasing age from x_0 .
ω	Oldest	$\log \mu_{x_1}$, the limiting mortality hazard at age x_1 and above.
δ	Time	Rate of change by calendar time of $\log(\text{mortality})$ at age x_0 .
λ_1	OBNRdecay	Rate of decay of OBNR working backwards in time from extract date.
ζ	SeasonalExcess	Amplitude of seasonal peak mortality from baseline (log scale); see equation (11).
τ	SeasonalPeak	Time of year of peak seasonal (winter) mortality, expressed as fraction of year from 1 st January; see equation (11).

B.3 Parameter correlations

A potential issue with the parametric model in Section 5 is that of confounding. Under normal circumstances, a parametric OBNR model will carry the risk of confounding with seasonal effects. For this reason we should include a seasonal term in an OBNR model, as per equation (11). However, the data extracts for the FRA and UK3 portfolios were also taken at a time of dramatic short-term changes in mortality levels caused by covid-19 [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020], so there remains potential for confounding even with a seasonal mortality model. Table 9 shows the percentage correlations between the parameters for the best-fitting OBNR model fitted to the FRA data set (this being the portfolio with the most pronounced OBNR effect). The OBNRdecay estimate, $\hat{\lambda}_1$, has a strong negative correlation with the time trend, δ , albeit it is not as strong as some other parameter pairings.

Table 9: Percentage correlations between parameter estimates in the model in Table 4 fitted to FRA data with an AIC of 164,008. Of note is the absence of any correlation between gender and either OBNR, season or time trend.

	AgeGradientYoungest	Gender.M	Gender.M:Oldest	Intercept	OBNRdecay	Oldest	SeasonalExcess	SeasonalPeak	Time
AgeGradientYoungest (m_0)	100	6	-12	-40	2	-63	0	-1	-10
Gender.M	6	100	-66	-13	0	31	0	0	0
Gender.M:Oldest	-12	-66	100	12	0	-42	0	0	0
Intercept (α)	-40	-13	12	100	37	17	-6	-11	-86
OBNRdecay (λ_1)	2	0	0	37	100	-9	-18	-16	-43
Oldest (ω)	-63	31	-42	17	-9	100	-2	1	8
SeasonalExcess (ζ)	0	0	0	-6	-18	-2	100	2	6
SeasonalPeak (τ)	-1	0	0	-11	-16	1	2	100	12
Time (δ)	-10	0	0	-86	-43	8	6	12	100

Table 10 shows the correlations between OBNRdecay and other parameter estimates when varying the length of the exposure period. The estimate and standard error of the OBNR decay parameter λ_1 are both robust to the length of the exposure period: from 1.3442 (s.e. 0.0532) with 4.5 years to 1.33501 (s.e. 0.0479) with 8.5 years of experience data (with $\hat{\lambda}_1 = 1.32332$ (s.e. 0.0508) using 5.5 years as in the main body of the paper).

Table 10: Variation by exposure-period term of percentage correlations between OBNRdecay ($\hat{\lambda}_1$) and other parameter estimates. FRA portfolio with June 2020 extract, Gaussian OBNR function from equation (8).

Exposure start	Years	AgeGradientYoungest	Gender.M	Gender.M:Oldest	Intercept	Oldest	SeasonalExcess	SeasonalPeak	Time
2016-01-01	4.5	2	0	0	43	-10	-19	-19	-47
2015-01-01	5.5	2	0	0	37	-9	-18	-16	-43
2014-01-01	6.5	2	0	0	32	-8	-17	-14	-39
2013-01-01	7.5	2	0	0	27	-8	-16	-13	-36
2012-01-01	8.5	2	0	0	24	-8	-15	-12	-33

Contact

More information including case studies, latest features, technical documentation and demonstration videos can be found on our website at www.longevitas.co.uk

24a Ainslie Place, Edinburgh, EH3 6AJ
Telephone 0131 315 4470
Email info@longevitas.co.uk

Longevitas is a registered trademark for Longevitas Ltd in the UK (registration number 2434941), throughout the European Union (registration number 5854518), and the USA (Trade Mark Registration No. 3707314).



LONGEVITASTM