DISCUSSION PAPER PI-2008

Modelling Seasonal Mortality with Individual Data

S.J. Richards, S.J. Ramonat, G.T. Vesper & T. Kleinow

May 2020

ISSN 1367-580X

The Pensions Institute
Cass Business School
City, University of London
106 Bunhill Row
London EC1Y 8TZ
UNITED KINGDOM
http://www.pensions-institute.org/

MODELLING SEASONAL MORTALITY

WITH INDIVIDUAL DATA

By S. J. Richards, S. J. Ramonat,
G. T. Vesper and T. Kleinow

LONGEVITAS™

Modelling seasonal mortality with individual data

Richards, Stephen J.*† Ramonat, Stefan J.‡ Vesper, Gregory T.§
Kleinow, Torsten¶

May 4, 2020

Abstract

Most studies of seasonal variation in mortality rely on aggregated death counts at population level. In this paper we use individual data to present a series of models for different aspects of seasonal variation. The models are fitted to a variety of international pensioner data sets and suggest a high degree of commonality across countries with different climates and different health systems. The power of individual life-history survival modelling allows the detection of seasonal patterns in even modest-sized portfolios. We measure the tendency for seasonal fluctuations to increase with age, and we again find strong similarities between geographically distinct populations. We further find that seasonal effects are generally uncorrelated with gender, but that low-income pensioners can suffer greater seasonal swings than high-income ones. Finally, we propose a single-parameter measure for the extent to which winter mortality is a spike and summer mortality is a shallower trough, and show results for a variety of data sets.

Keywords: seasonal mortality, excess winter mortality, survival model.

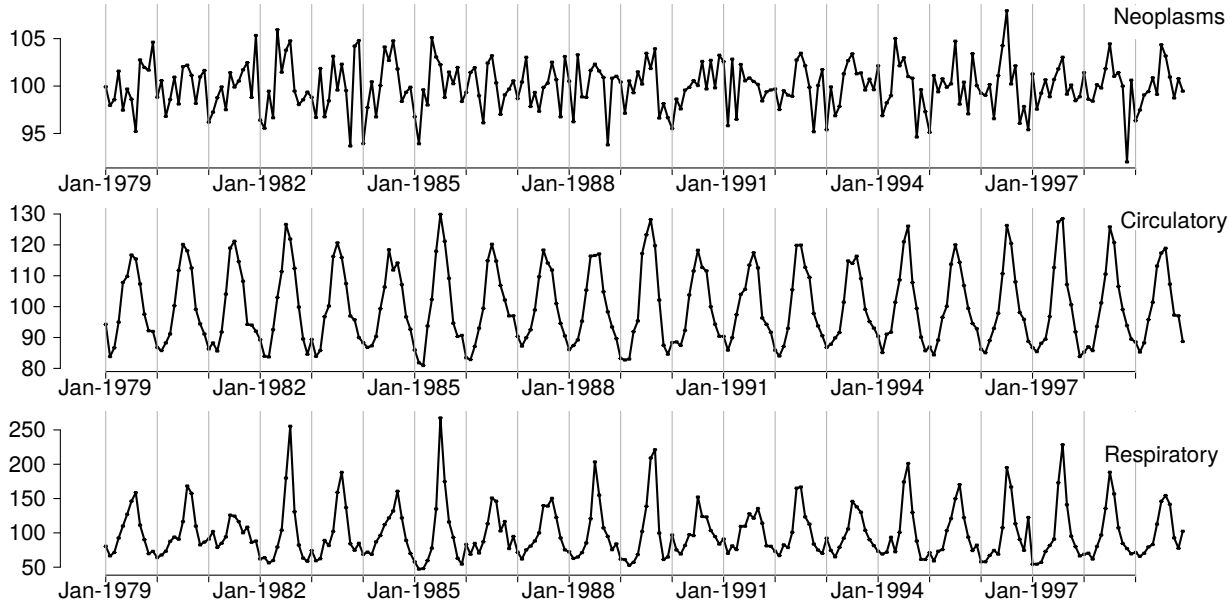
1 Introduction

It has long been known that human mortality levels vary by season — Figure 1 demonstrates the durability of this pattern for selected causes of death in Australia, a modern, developed urban nation with a good healthcare system. We will take the seasonality of mortality as read, but interested readers can consult Rau [2007, Chapter 2] for a comprehensive introduction to the topic, covering both historical and modern aspects. From a medical perspective, Boulay et al. [1999] discuss the physiological aspects and possible causal pathways for chronic heart failure, a cause of death with pronounced seasonality.

Most approaches to measuring seasonal mortality involve observed death counts at the population level [Rau, 2007, Chapter 3]. This is due to the large volumes of data that are required to detect patterns for intervals of time much shorter than a year. In this paper we present an alternative approach based on the mortality of individuals using survival models [Macdonald et al., 2018]. We present a series of models for various aspects of seasonal mortality and illustrate them using data sets of retirees receiving income from occupational pension schemes or insurer annuities. Although these data sets are orders of magnitude smaller than national populations, the power of modelling the mortality of individuals, rather than groups, means that seasonal effects are measurable for even modest-sized pension schemes. A further benefit of using individual data is the rich variety of additional covariates that becomes available [Richards, 2020, Section 8]. In contrast, grouped data

*Corresponding author: stephen@longevitas.co.uk
†Longevitas Ltd, Edinburgh. www.longevitas.co.uk. http://orcid.org/0000-0001-6859-6347
‡Mercer, 55 Metcalfe Street, Suite 550, Ottawa, Ontario K1P 6L5, Canada. https://orcid.org/0000-0003-0366-2942
§gvesper@gmail.com
¶Heriot-Watt University, Edinburgh, EH14 4AS, and Maxwell Institute for Mathematical Sciences, James Clerk Maxwell Building, Mayfield Road, Edinburgh EH9 3FD. https://orcid.org/0000-0001-8027-1430

Figure 1: Percentage of average daily number of deaths for selected causes in Australia, 1979–1999. Source: de Looper [2002]. Over the period concerned the population of Australia grew from 14 million to 18 million.



have to be stratified with respect to such factors, which can severely limit what can be achieved [Macdonald et al., 2018, Chapter 7.3].

Section 2 describes the data sets used in this paper and Section 3 details the basic form of the corresponding survival model. Section 4 introduces a simple approach to modelling seasonal mortality and presents results for a wide variety of pension schemes in different countries. Section 5 considers modelling seasonal variation by age. Section 6 looks at modelling the shape of seasonal variation, in particular the tendency for mortality to spike in winter, as opposed to a relatively shallow trough in summer. Section 7 examines variation by sub-groups, such as gender and income level. Section 8 considers applications, while Section 9 concludes.

2 Data

The data sets used in this paper consist of pensioner records from occupational pension schemes, insurer annuity portfolios or national pension systems. Due to the financial interest in not paying pensions longer than necessary, such portfolios usually maintain accurate records of when pensions commence and when they cease. Pension schemes are specific to a particular employer, and so pensioners often share an occupational background and in some cases are geographically concentrated. Due to regulations and tax-reporting requirements, pension schemes often have detailed additional information besides date of birth, gender and pension. Such portfolios are like longitudinal studies with continuous recruitment: as people retire, new benefit records are set up. Upon the death of a former employee, a surviving spouse's pension might also be set up. Table 1 gives an overview of the portfolios.

Figure 2: Seasonality of deaths for data set UK2 in Table 1 with least-squares fit of cosine shape. The vertical scale excludes an outlier caused by leap years.

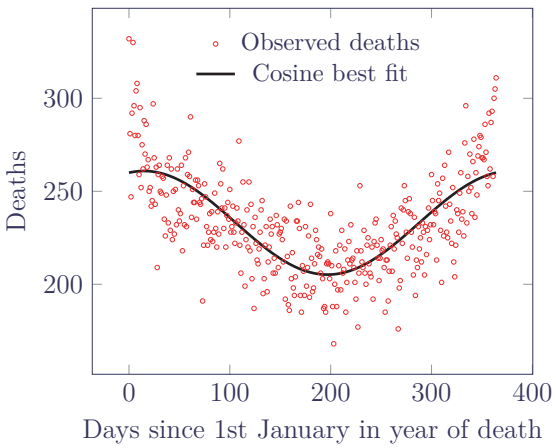


Table 1: Overview of data in portfolios. The exposure periods vary within the range 1982–2018, i.e. excluding the period of the COVID-19 pandemic [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020].

Country	Portfolio	Age range	Deaths	Description
Canada	CANEAST	55–100	20,790	Members of defined-benefit occupational schemes with known addresses in Québec and the Atlantic Provinces (NL, NB, NS, PE).
	CANWEST	55–100	36,047	Members of defined-benefit occupational schemes with known Canadian addresses outside Québec and the Atlantic Provinces.
	CAN1	55–100	57,792	Combination of CANEAST and CANWEST above plus small number with overseas or unknown addresses.
	CAN2	55–100	2,614	Single-employer defined-benefit occupational pension scheme in CAN1.
Chile	CHL	55–95	179,987	Pensioner records from Superintendencia de Pensiones for 1982–2013.
England	ENG	60–105	19,435	Defined-benefit occupational pension scheme for a single English local authority.
Spain	ESP	55–104	3,822	Some annuities, but mainly recipients of income from investment bonds. Higher-net-worth individuals concentrated in southern Spanish mainland.
France	FRA	55–99	35,665	Insurer portfolio of voluntary top-up pensions for employees of higher-education institutions around France.
Kuwait	KUW	50–76	10,178	Kuwaiti nationals who are contributors, pensioners or beneficiaries under the social security programs. The age range is restricted as birth registration in Kuwait only began in 1952; see Hill [1975].
Netherlands	NLD	50–105	4,896	Single-employer occupational pension scheme in the private sector.
Scotland	SCOT	50–105	3,488	Defined-benefit occupational pension scheme for a single Scottish local authority, as used in Richards [2020].
UK	UK1	60–95	58,213	Individual annuitants from defined-contribution personal pensions, widely spread around the UK.
	UK2	60–104	60,661	Defined-benefit occupational pension schemes for six local authorities, including ENG and SCOT portfolios above.
USA	USA1	65–103	32,518	Single-employer defined-benefit occupational pension scheme for a non-union, blue-collar workforce concentrated in one geographical region, but with pension-holders across the US. Dates were recorded on a year-and-month basis only, resulting in some imprecision in the seasonal parameters.
	USA2	55–100	10,594	Occupational pension scheme in the US.

An important difference between these data sets and other work on seasonal mortality is that these are individual records, not grouped counts. Also, no cause-of-death codes are available, so we are dealing with all-cause mortality, as opposed to the respiratory-only deaths of Eilers et al. [2008] or the cerebrovascular and respiratory deaths of Marx et al. [2010]. Since these specific causes exhibit more pronounced seasonal variation (see Figure 1), our all-cause mortality will exhibit more muted seasonal patterns. Nevertheless, as shown in Figure 2, seasonal variation is still evident.

Another important aspect of individual records is the potential presence of duplicates, i.e. two or more pensions or annuities paid to the same person. For portfolios CANEAST, CANWEST, CAN1, ENG, ESP, KUW, SCOT, UK1 and UK2 in Table 1 the records have been deduplicated using the techniques described in Macdonald et al. [2018, Section 2.5]. Individual records also enable detailed data-quality checks to be carried out — see Macdonald et al. [2018, Sections 2.3, 2.4, 2.7 and 2.8]. To avoid risk of age mis-statements distorting model fits — see Newman [2018a] and Newman [2018b] — any pensioner appearing to exceed age 105 was excluded from the data set. The use of individual data makes data-quality checking far easier than with grouped counts. For example, the FRA data set had nearly 700 annuitants appearing to reach age 110; however, since each shared the same date of birth, it was clear that these records were erroneous.

3 Basic model and fitting

To fit a survival model to individual data we maximise the log-likelihood function in equation (1) [Macdonald et al., 2018, Section 5.3] for the mortality hazard, $\mu_{x,y}$, at age x at calendar time y. 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 observed to die at age x_i+t_i, or 0 otherwise. $H_{x,y}(t)$ is the integrated hazard function in equation (2). Appendix A contains the details of how we performed the numerical integration of the periodic functions underlying the models.

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

$$H_{x,y}(t) = \int_0^t \mu_{x+s,y+s} ds \quad (2)$$

An important point to note about equation (1) is that it is the log-likelihood for a survival model with left-truncated data, which is standard for data encountered in actuarial work; see Macdonald et al. [2018, Section 4.3]. This contrasts with survival models used in medical research, where left-truncation is relatively uncommon and where likelihoods are usually for non-left-truncated data [Collett, 2003, Chapter 6].

4 Basic seasonal variation

We assume a suitable function for the age- and time-varying mortality hazard, $\mu_{x,y}$, is available. The choice will depend on the age range under study, but here we will use the Hermite II model of Richards [2020] for its particular suitability for mortality rates at post-retirement ages. We start with the cosine extension for seasonal mortality of Richards [2020, Section 8], as shown in equation (3). $\tau \in [0, 1)$ 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).

$$\log \mu^*_{x,y} = \log \mu_{x,y} + e^{\zeta} \cos\left(2\pi(y - \tau)\right) \quad (3)$$

The two-parameter model for seasonal mortality in equation (3) seeks to simultaneously identify (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 $2e^{\zeta}$. This contrasts with the two-parameter model of Gemmell et al. [2000], who used both cosine and sine

functions and where the amplitude and peak were derived from the model parameters; Eilers et al. [2008, Section 3] similarly used a combination of sine and cosine functions. However, here the definition of equation (3) 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. Unlike Eilers et al. [2008, Section 3] we do not attempt to simultaneously estimate a time-trend component, although this is possible with the Hermite models; see Richards [2020, Section 7]. Equation (3) specifies an addition to $\log \mu_{x,y}$, so the mortality hazard is multiplied by a factor fluctuating between $\exp(-e^{\zeta})$ and $\exp(e^{\zeta})$.

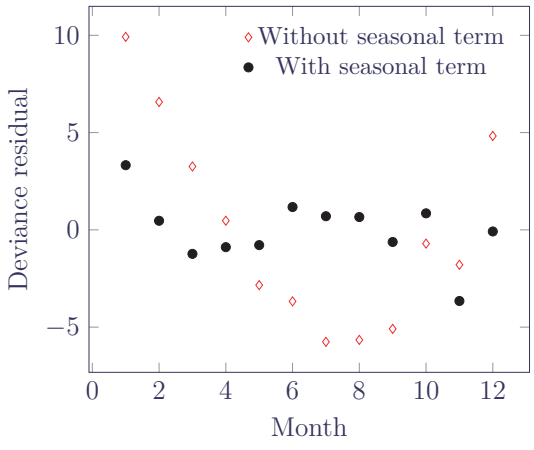
Parameters are estimated by maximising the log-likelihood in equation (1) and Table 2 shows the values of $\hat{\zeta}$ and $\hat{\tau}$ for a variety of pensioner populations. The high degree of agreement across a wide geographic range suggests a common underlying mechanism — potential physiological aspects are discussed in Boulay et al. [1999]. Kloner et al. [1999] suggested that the seasonal peak in deaths due to coronary artery disease in Los Angeles County might be partly due to "factors such as overindulgence or the stress of the holidays", referring to the Christmas & New Year period. However, as the inhabitants of a majority-Muslim country, few Kuwaitis will be drinking much alcohol or feasting in celebration of Western Christian holidays, yet they have the same seasonality to their mortality. Furthermore, Chileans celebrate Christmas at the same time as the northern-hemisphere countries, and yet their mortality peak occurs six months later. Whatever underlies the common seasonal pattern in Table 2, it is unlikely to be excessive alcohol consumption or Christmas over-eating. Table 2 shows that the peak is always during the winter: January or February in the northern hemisphere and July in the southern hemisphere.

Table 2: $\hat{\zeta}$ and $\hat{\tau}$ for pensioner groups in portfolios from Table 1. Source: own calculations.

Portfolio	Excess $\hat{\zeta}$	Peak $\hat{\tau}$	Peak mortality: (i) as % of mean	(ii) time of year
CANEAST	-2.06	0.0976	114%	6th Feb
CANWEST	-2.47	0.0538	109%	20th Jan
CAN1	-2.34	0.0749	110%	27th Jan
CAN2	-2.04	0.1090	114%	9th Feb
CHL	-2.25	0.5560	111%	22nd Jul
ENG	-2.02	0.0708	114%	26th Jan
ESP	-2.89	0.1494	106%	24th Feb
FRA	-2.42	0.0660	109%	25th Jan
KUW	-2.36	0.0178	110%	7th Jan
NLD	-2.25	0.0524	111%	20th Jan
SCOT	-1.88	0.0815	117%	30th Jan
UK1	-2.28	0.0885	111%	2nd Feb
UK2	-1.95	0.0713	115%	27th Jan
USA1	-2.52	0.1420	108%	21st Feb
USA2	-2.63	0.0739	107%	27th Jan

How effective can the simple cosine seasonal term in equation (3) be, given the excess deaths above the peak in Figure 2? Figure 3 shows that the improvement in fit is quite marked: the Poisson deviance residuals [McCullagh and Nelder, 1989] for the CAN1 data set are much smaller in magnitude and the pattern is much reduced after including even just a simplistic cosine term. Appendix C contains further comparisons of residual patterns with and without seasonal terms for the CHL, FRA and UK2 portfolios.

Figure 3: Deviance residuals for CAN1 data set with and without seasonal term in equation (3).



5 Seasonal variation by age

Another known facet of seasonal variation in mortality is that it increases with age; see Rau [2007, pages 33–36, Figures 2.14 and 2.15] and Richards [2020, Figure 13]. We can model this by extending equation (3) to include an age-dependent seasonal factor as in equation (4):

$$\log \mu^*_{x,y} = \log \mu_{x,y} + e^{\zeta+\xi(x-o)/10} \cos\left(2\pi(y-\tau)\right) \tag{4}$$

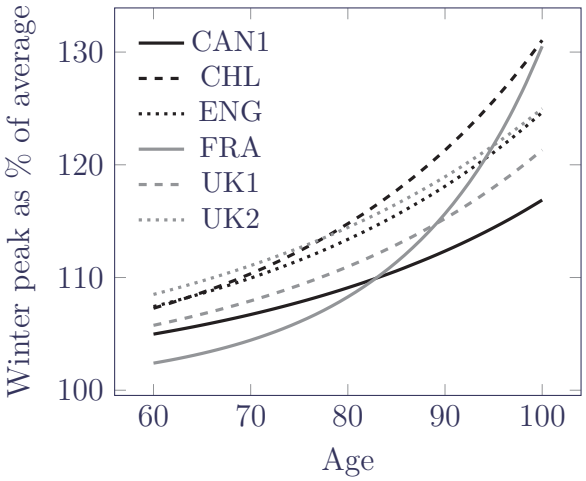
where the new parameter, ξ, measures the extent to which seasonal variation changes with age x. The factor of 1/10 is to keep the parameters well-scaled. The age offset, o, reduces the correlation between ζ and ξ that otherwise causes problems in optimising the log-likelihood; in this paper we use $o = 70$. Appendix B contains an overview of all the parameters used in this paper.

Since seasonal fluctuations tend to increase with age, we expect a positive value for $\hat{\xi}$. Table 3 shows the estimated parameters for the five largest portfolios, and Figure 4 shows the resulting modelled peak winter mortality as a percentage of the average. There is a high degree of commonality among CAN1, CHL, ENG, UK1 and UK2, but the FRA data set has a noticeably steeper rate of increase with age.

Table 3: Seasonal parameter estimates for equation (4) for selected portfolios from Table 1. Source: own calculations with $o = 70$.

Portfolio	Excess $\hat{\zeta}$	Age $\hat{\xi}$	Peak $\hat{\tau}$
CAN1	-2.73	0.291	0.0768
CHL	-2.32	0.337	0.5562
ENG	-2.36	0.281	0.0671
FRA	-3.14	0.604	0.0626
UK1	-2.57	0.310	0.0845
UK2	-2.25	0.251	0.0703

Figure 4: Modelled peak winter mortality by age as percentage of average. Source: own calculations of $\exp(e^{\zeta}) \times 100\%$ from parameter estimates in Table 3.



6 Shape of seasonal variation

In their review of deaths due to coronary artery disease in Los Angeles County, Kloner et al. [1999] noted a quadratic U-shape to the pattern of deaths throughout the calendar months. This is echoed in Figure 2, but note that the pattern is shifted by six months in the southern hemisphere, as shown in Figure 1. Similarly, Marx et al. [2010] noted "sharp peaks in winter and relatively flat troughs in summer". Clearly, seasonal mortality variation has a spikier nature in winter and a shallower, trough-like shape in summer. We could allow for this using polynomial splines: one for winter and one for summer, with two blending points a quarter year before and after the winter peak (say). These splines would be functions of two parameters, one measuring the spikiness of the winter peak and one measuring the shallowness of the summer trough. However, since both peak and trough can be expected to move in tandem, ideally we would have a single parameter to achieve both effects simultaneously.

One way to sharpen winter peaks while simultaneously flattening summer troughs is shown in equations (5) and (6). We have a single parameter, ψ, representing the shape of the seasonal pattern within the year, and which we can estimate from the data. This contrasts with Marx et al. [2010, Section 2.2], who used individual parameters for each month.

$$\log \mu^*_{x,y} = \log \mu_{x,y} + e^{\zeta+\xi(x-o)/10} s(2\pi(y-\tau)) \quad (5)$$

$$s(t) = \begin{cases} \psi \neq 0 & : 2\left[\dfrac{e^{\frac{\psi}{2}(1+\cos t)}-1}{e^{\psi}-1}\right] - 1 \\ \psi = 0 & : \cos t \end{cases} \quad (6)$$

Figure 5: Shape of seasonal effect, $s(t)$, in equation (6) with different values of ψ.

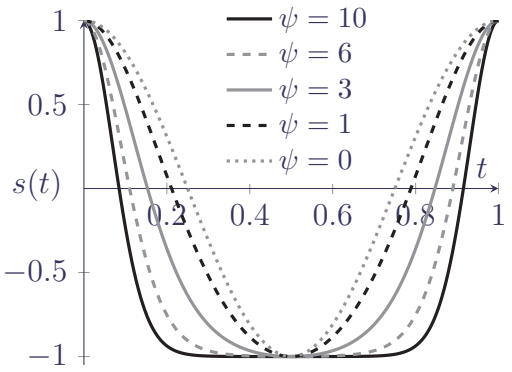


Figure 5 shows the range of shapes produced by different values of ψ in equations (5) and (6). Note that if ψ should not prove to be significantly different from zero, then one could revert to equation (3) or (4). Also, while it is unlikely to ever arise in practice, in theory ψ could take a negative value; this would mean that the summer trough was sharper than the winter peak. Equations (5) and (6) are by no means the only possible single-parameter approach — Appendix D describes an alternative formula that behaves in a similar manner.

Tables 4 and 5 show the parameter estimates for several portfolios, where each country demonstrates a strong degree of peakedness for winter mortality, i.e. $\psi >> 0$; of particular note is the very flat summer trough of the KUW portfolio. A comparison of $\hat{\tau}$ and $\hat{\psi}$ between Tables 4 and 5 reveals that estimates for these parameters are little affected by the presence or absence of the age-related parameter, ξ, in the model. Checks were made for CAN1, FRA and UK2, but no significant variation in shape by age was found. However, this does not mean that no such variation exists: shape differences are secondary in strength to age differences (see Table 7) and can only be detected with either larger portfolios (CAN1, CHL, ENG, FRA, UK1 and UK2) or where the shape is particularly extreme (KUW).

Table 4: Seasonal parameter estimates for equation (5) with $\xi = 0$ for selected portfolios from Table 1. Source: own calculations.

Portfolio	Excess $\hat{\zeta}$	Peak $\hat{\tau}$	Shape $\hat{\psi}$	Signature
CAN1	-2.31	0.0719	2.11	
CHL	-2.23	0.5464	1.93	
ENG	-2.00	0.0573	2.41	
FRA	-2.38	0.0662	2.28	
KUW	-2.16	0.0105	6.02	
UK1	-2.26	0.0638	2.37	
UK2	-1.95	0.0668	1.02	

Table 5: Seasonal parameter estimates for equation (5) for selected portfolios from Table 1. Source: own calculations with $o = 70$.

Portfolio	Excess $\hat{\zeta}$	Age $\hat{\xi}$	Peak $\hat{\tau}$	Shape $\hat{\psi}$	Signature
CAN1	-2.66	0.262	0.0753	1.98	
CHL	-2.29	0.311	0.5486	1.76	
ENG	-2.29	0.252	0.0565	2.21	
FRA	-3.06	0.576	0.0644	2.04	
UK1	-2.58	0.340	0.0625	2.41	
UK2	-2.25	0.247	0.0661	1.00	

7 Seasonal variation by sub-groups

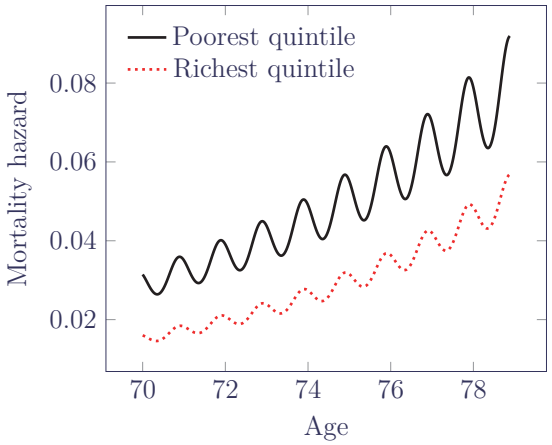
Table 2 shows a high degree of commonality across different countries, albeit with the peak for Chile shifted by six months for the winter in the souther hemisphere. It also shows that seasonal patterns are so strong that they can be detected in a portfolio with fewer than 3,000 deaths, provided the exposure period covers multiple years. In this section we consider variation by sub-groups of each portfolio.

Section 5 showed that seasonality varies by age, but what of the other classic risk factors of mortality, such as gender and income level? Such investigations are a challenge for grouped data due to the problem of stratification: as one partitions the data set into smaller parts, the statistical procedures lose power when applied to each part separately. [Macdonald et al., 2018, Table 7.1] shows how even a large data set quickly loses its ability to support a multi-factor model. In contrast, the ability to use covariates is a particular strength of modelling mortality at the level of the individual.

We found that parameters for seasonal effects are largely the same for men and women. For example, there was no significant difference between seasonal excess mortality for males and females in the CAN1, FRA, KUW, NLD, SCOT, UK1 and UK2 datasets, a result echoed by Gemmell et al. [2000], who found only a moderately higher male susceptibility to seasonal variation for males in Scotland. Even for the CHL dataset, with over half a million male lives and over 700,000 females, we found no difference in seasonal excess between men and women. The CAN1 dataset was the only one to come close to finding a significant interaction between seasonal excess (ζ) and gender, with a p-value of 0.065.

The picture is different for pension income, however. The FRA, NLD and UK1 datasets saw only a mild association between seasonal variation and pension income, and CAN1 not at all. In contrast, for the UK2 dataset those with incomes in the lowest quintile had statistically significantly higher susceptibility in winter. Figure 6 shows the extent of modelled mortality differentials for males in the poorest and highest income quintiles; not only do the poorest have the highest mortality, but they also have the highest seasonal variation as well. This echoes Gemmell et al. [2000, Figure 2], who found that Social Classes I and II had a lower peak winter excess than the more-deprived Social Classes III-V. The most deprived social classes tend to be under-represented in UK pension schemes, making it all the more remarkable that the link between seasonal variation and income could be detected in a data set that is unrepresentative of the wider national population. Note that the pension-income effect is not a proxy result for the tendency for older pensioners to have smaller pensions, as the seasonal age parameter, ξ, was also present in the model. Continuous Mortality Investigation [2019, page 16] noted that the most-deprived deciles of the UK population have a cause-of-death mix skewed towards causes exhibiting the highest degree of seasonal variation.

Figure 6: Seasonality of mortality hazard by income quintile for data set UK2. Source: own calculations of modelled rates for males only.



Why is different seasonality by income level only evident in the UK2 data set, and not in others? We believe that the difference lies in the nature of the data sets, specifically their ability to indicate the likely total retirement income. For example, the FRA data set is for top-up pensions where the median annual amount was just €1,335 — the stated pension amounts are therefore not a

wholly reliable guide to total retirement income. Similarly, the UK1 data set is of annuitants from individual defined-contribution pension savings plans — such annuitants will often have other sources of retirement income, so again the pension amount is unlikely to be a reliable guide to total retirement income. Amongst the rather diverse schemes underlying CAN1, mortality varies only modestly by pension amount; part of the reason might again be due to the pension-plan data only giving an incomplete picture of total retirement income. For example, Canadian employers might pay top-up pensions in excess of the limits on tax-deferred pension amounts directly from corporate funds, which might therefore not be recorded in the pension-administration system. In contrast, the UK2 data set is for defined-benefit occupational pensions in the public sector — here the pension is more likely to be the main source of retirement income, and so is a more reliable guide to total income.

8 Applications

The primary application of seasonal modelling is perhaps for social policy and public health — Figure 6 shows that there is potentially avoidable excess winter mortality amongst the poorest in the UK. In contrast, seasonal variation in mortality is normally of minor actuarial interest: the cyclic nature of the pattern means that the impact on reserves for long-term business like pensions and annuities is usually small, despite the strong statistical significance [Richards, 2020, Section 9]. However, there are three operational applications that make seasonal modelling of use to actuaries in life insurance or involved in the analysis of pensioner mortality.

The first actuarial application lies in checking data quality. Macdonald et al. [2018, Sections 2.3–2.7] describes the advantages of using individual records, and analysis of seasonal patterns — or their absence — is a useful additional check of data validity. For example, the estimated shape parameter for UK2 in Tables 4 and 5 is out of line with the values for UK1, CAN1 and FRA. Further inspection highlighted some suspicious heaping of dates of death, possibly indicating that these were dates of bulk updates, rather than actual dates of death. This sort of data anomaly is something that a writer of a bulk annuity or longevity swap would want to know about before committing to a pricing basis.

A second actuarial application lies in analysing recent trends, where allowance for seasonal variation is important to minimise any bias in trend estimation from the choice of exposure period. Richards [2020, Section 8] cites the example of an insurer looking to analyse the mortality experience of a pension scheme or other portfolio prior to a bulk-annuity transaction or longevity swap. In such circumstances, insurers need to use all available information, and stretches of the exposure period should not have to be discarded just to even up the representation of the seasons. The need to consider seasonal mortality in analysing mortality-improvement rates was specifically highlighted by the UK insurance regulator in a letter to chief actuaries; see Malik [2019]. The COVID-19 pandemic [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020] may create a more general need to model mortality levels in the presence of a mortality spike, especially if it becomes one of the recurrent seasonal respiratory viruses [Monto et al., 2020].

A third actuarial application is operational risk and staffing levels for administering large portfolios, especially rapidly growing ones. For example, an insurer in the bulk-annuity market might take on a particularly large new liability with tens of thousands of extra lives — the use of a model allowing for the shape of the seasonal mortality variation in Section 6 permits month-by-month planning of staffing levels in administration areas. Of particular note is the tendency for large transactions to be concluded at the end of the calendar year, i.e. a month prior to the typical surge in winter mortality shown in Table 2 for countries in the northern hemisphere.

9 Conclusions

Survival models with individual data can detect seasonal variation by age amongst retired people, even in modest-sized pension schemes. The seasonal peak in winter mortality is consistent across a wide range of territories and populations, with winter mortality typically peaking at 106–117% (late January and early February in the northern hemisphere, July in the southern). Seasonal patterns are sufficiently strong that survival models can reliably detect the increase in seasonal variation with age, which is also consistent across different countries: modelled peak winter mortality rises from 103–108% at age 60 to around 115-130% at age 100. Individual survival models can also detect the different shape of the winter mortality spike, compared to the shallower summer mortality trough. The parameters describing seasonal variation are largely uncorrelated with other mortality factors, such as gender, but there is evidence that seasonal variation is more extreme in the UK for those with the smallest incomes.

Acknowledgments

The authors thank the following for their support in providing data: Monzer Mourad of the Public Institution for Social Security (PIFSS) of Kuwait, who graciously provided the data analysed; Wendy Ho of the AIHW for the data on seasonal mortality in Australia; Hélène Queau and Clément Frappier; Alex Oleksandr Zavershynskyy; Denis Dupont; Jessica Hartman, Allison Steinmetz and Claire y Li; Jay Wang and Jonathan Merl; and Andres Barajas Paz. The authors also thank Gavin Ritchie and Angus Macdonald for helpful comments on earlier drafts of this paper. Any errors or omissions remain the responsibility of the authors. Models were fitted using the Longevitas survival-modelling system and bespoke programs written in C++ and R [R Core Team, 2017]. Graphs were produced in R and tikz, while typesetting was done in LaTeX.

References

H. Akaike. Factor analysis and AIC. *Psychometrica*, 52:317–333, 1987. ISSN 0033–3123. doi: https://doi.org/10.1007/BF02294359.

F. Boulay, F. Berthier, O. Sisteron, Y. Gendreike, and P. Gibelin. Seasonal variation in chronic heart failure hospitalizations and mortality in france. *Circulation*, 100:280–286, 1999.

D. R. Butenhof. *Programming with POSIX Threads*. Addison-Wesley, 1997. ISBN 978-0-201-63392-4.

D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall, second edition, 2003. ISBN 1-58488-325-1.

Continuous Mortality Investigation. *CMI Mortality Projections Model: Interim update*. Working Paper 127, 2019.

M. de Looper. *Seasonality of death*, volume Bulletin No. 3. Australian Institute of Health and Welfare, 2002. ISBN 978–1–74024–209–7.

P. H. C. Eilers, J. Gampe, B. D. Marx, and R. Rau. Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, 27(17):3430–3441, 2008. doi: 10.1002/sim.3188.

I. Gemmell, P. McLoone, F. A. Boddy, G. J. Dickinson, and G. C. M. Watt. Seasonal variation in mortality in Scotland. *International Journal of Epidemiology*, 29:274–279, 2000.

A. G. Hill. The Demography of the Kuwaiti Population of Kuwait. *Demography*, 12(3):537–548, 1975. ISSN 0070-3370. doi: 10.2307/2060834.

William Kahan. Further remarks on reducing truncation errors. *Communications of the ACM*, 8(I): 40, 1965.

R. A. Kloner, W. K. Poole, and R. L. Perritt. When throughout the year is coronary death most likely to occur? A 12-year population-based analysis of more than 220,000 cases. *Circulation*, pages 1630–1634, 1999.

A. S. Macdonald, S. J. Richards, and I. D. Currie. *Modelling Mortality with Actuarial Applications*. Cambridge University Press, 2018. ISBN 978-1-107-04541-5.

S. Malik. Observations from recent regulatory reviews. Letter to Chief Actuaries of UK-regulated insurers and reinsurers, June 2019. URL `https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/letter/2019/observations-from-recent-regulatory-reviews`.

B. D. Marx, P. H. C. Eilers, J. Gampe, and R. Rau. Bilinear modulation models for seasonal tables of counts. *Statistics and Computing*, 20(2):191–202, 2010. doi: 10.1007/s11222-009-9144-9.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, second edition edition, 1989. ISBN 0-412-31760-5.

A. S. Monto, P. DeJonge, A. P. Callear, L. A. Bazzi, S. Capriola, R. E. Malosh, E. T. Martin, and J. G. Petrie. Coronavirus occurrence and transmission over 8 years in the HIVE cohort of households in Michigan. *The Journal of Infectious Diseases*, 04 2020. ISSN 0022-1899. doi: 10.1093/infdis/jiaa161. URL `https://doi.org/10.1093/infdis/jiaa161`. jiaa161.

S. J. Newman. Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biology*, 16(12):e2006776, 2018a. doi: https://doi.org/10.1371/journal.pbio.2006776.

S. J. Newman. Plane inclinations: A critique of hypothesis and model choice in Barbi et al. *PLoS Biology*, 16(12):e3000048, 2018b. doi: https://doi.org/10.1371/journal.pbio.3000048.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

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.

Superintendencia de Pensiones. Tablas de Mortalidad en Formato. https://www.spensiones.cl/apps/bdp/index.php.

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`.

An interactive online tool to explore the functional forms presented in this paper can be found at https://www.longevitas.co.uk/site/Hermite/Season.html

Appendices

A Implementation

A general approach to fitting survival models is presented in Richards [2020, Appendix B]. The starting point is the log-likelihood, ℓ, defined in equation (1) and the contribution of a single life i is given in equation (7). The first partial derivative of equation (7) with respect to a parameter θ is as follows:

$$\ell_i = -\left[\int_0^{t_i} \mu_{x_i+s,y_i+s} ds\right] + d_i \log \mu_{x_i+t_i,y_i+t_i} \quad (7)$$

$$\frac{\partial}{\partial\theta}\ell_i = -\left[\int_0^{t_i} \mu_{x_i+s,y_i+s} \left(\frac{\partial}{\partial\theta} \log \mu_{x_i+s,y_i+s}\right) ds\right] + d_i \frac{\partial}{\partial\theta} \log \mu_{x_i+t_i,y_i+t_i} \quad (8)$$

and the second partial derivative with respect to a pair of parameters, θ_1 and θ_2, is as follows:

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ell_i = -\left[\int_0^{t_i} \mu_{x_i+s,y_i+s} \left[\frac{\partial^2}{\partial\theta_1\partial\theta_2} \log \mu_{x_i+s,y_i+s} + \left(\frac{\partial}{\partial\theta_1} \log \mu_{x_i+s,y_i+s}\right)\left(\frac{\partial}{\partial\theta_2} \log \mu_{x_i+s,y_i+s}\right)\right] ds\right] + d_i \frac{\partial^2}{\partial\theta_1\partial\theta_2} \log \mu_{x_i+t_i,y_i+t_i} \quad (9)$$

In each of equations (7)-(9) we have integrands that fluctuate due to the seasonal component — Figure 7 shows an example.

Figure 7: Integrands for equations (7)-(9). Source: own calculations for hazard for a male in the poorest quintile in UK2 data set in Figure 6.

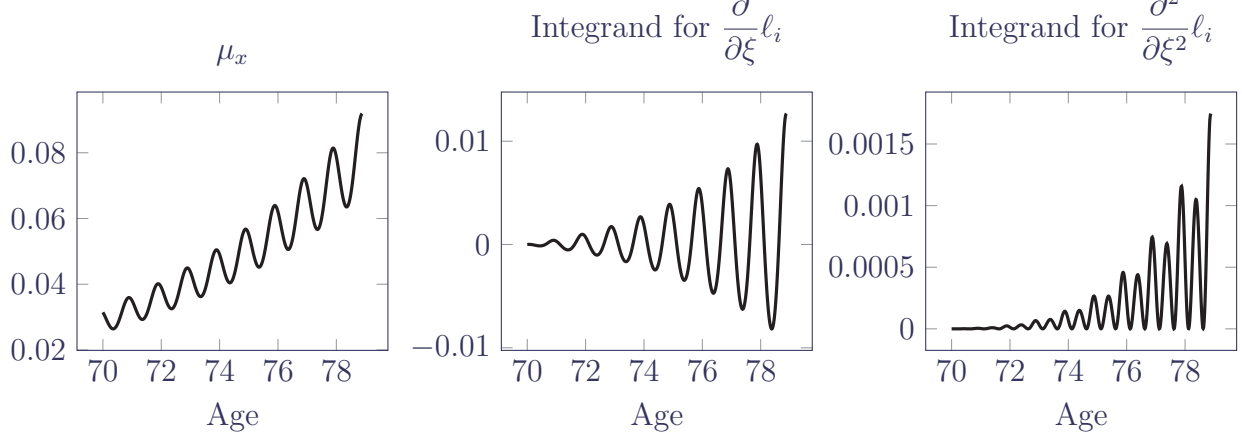
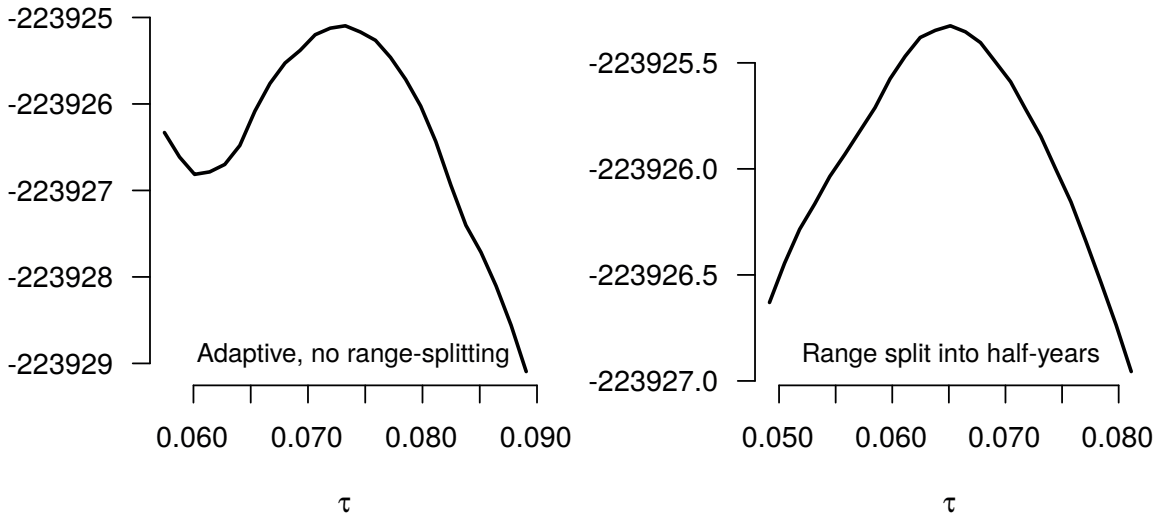


Figure 7 illustrates why fitting survival models with seasonally varying hazard functions is tricky. For integrating the mortality hazard (left) and the integrand for any first partial derivative (centre), we achieve best results by splitting the region of integration into intervals of at most half a year, aligning on the anniversaries of τ and $\tau + 0.5$ in each calendar year. For the integrand of the second partial derivative, the right-hand panel of Figure 7 shows that there are actually two complete cycles within each calendar year, and so for calculating the information matrix it is best to split the region of integration into quarter-year intervals aligning on the anniversaries of τ, $\tau + 0.25$, $\tau + 0.5$ and $\tau + 0.75$. Splitting the region of integration has a computation cost when repeated for hundreds of thousands of lives, so we used parallel processing over 63 threads to reduce run-times [Butenhof,

1997]. Splitting the region of integration also adds to the risk of truncation errors when adding up millions of small values, so we use the algorithm of Kahan [1965] to mitigate this when evaluating the likes of equations (7), (8) and (9).

For the results in this paper we have either used Romberg integration, Clenshaw-Curtis integration or Gaussian quadrature. A useful test for whether the approximation of integrals was successful or not is to plot the profile log-likelihood for the parameters. We found that failure to have a cleanly inverted-U shape was often a sign that the numerical integration was insufficiently rigorous — see Figure 8 for an example. Note that the converse is not necessarily true — a cleanly inverted U-shape for the profile log-likelihood is not on its own sufficient evidence that the integration was accurately performed.

Figure 8: Profile log-likelihoods of of mortality hazards with seasonal effects. Left: adaptive algorithm that otherwise works well for non-fluctuating hazards. Right: algorithm applied between each seasonal peak and trough. Source: own calculations.



B Parameter overview

Table 6 sets out the parameters whose values are estimated from the data.

Table 6: Overview of seasonal parameters.

Parameter	Name	Description and role of parameter
ξ	SeasonalAge	Rate of increase by age of peak seasonal mortality; see Section 5.
ζ	SeasonalExcess	Amplitude of seasonal peak mortality from baseline (log scale); see Section 4.
τ	SeasonalPeak	Time of year of peak seasonal (winter) mortality, expressed as fraction of year from 1st January.
ψ	SeasonalShape	Shape of winter peak and summer trough, with $\psi = 0$ implying equal curvature of winter peak and summer trough; see Section 6.

C Model fit and further research

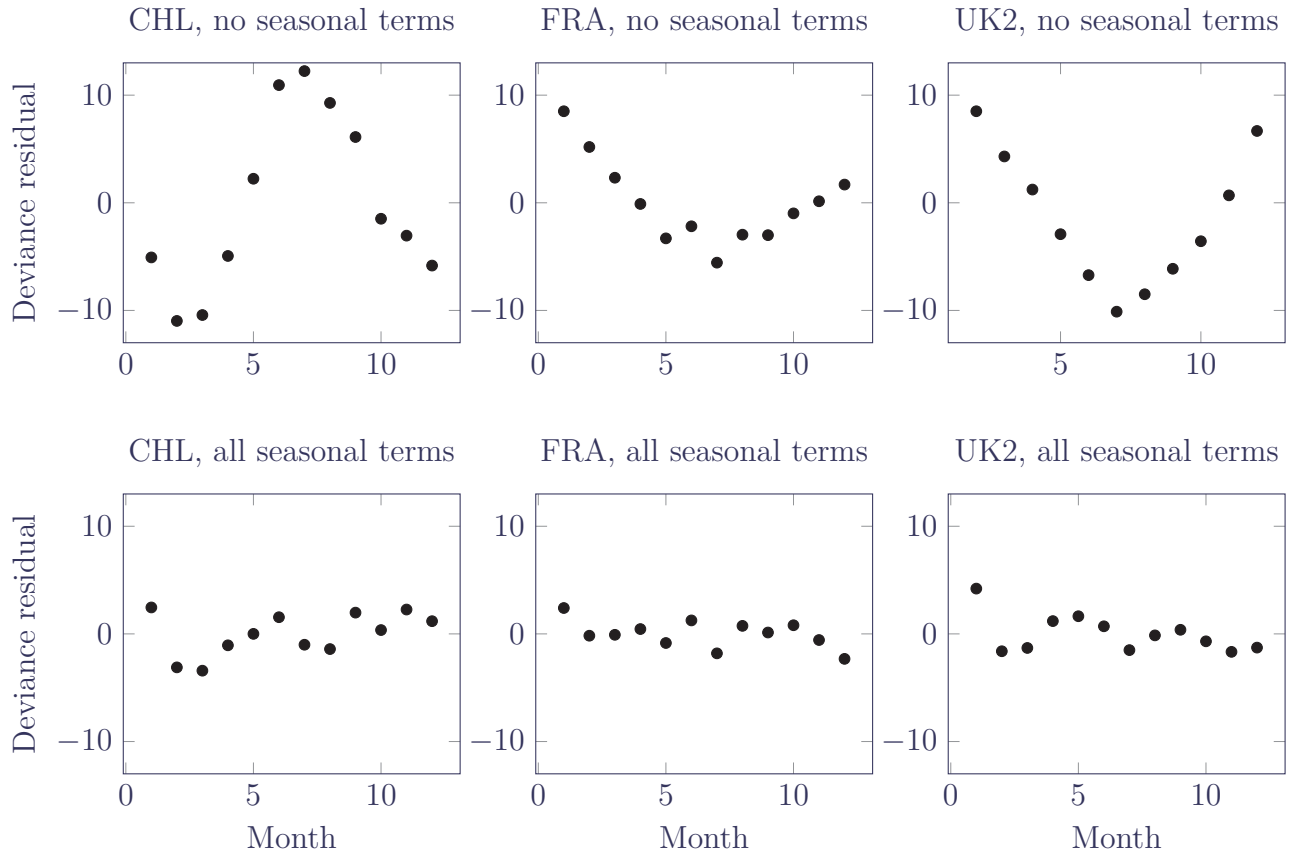
Table 7 shows the improvements in model fit achieved from adding the various seasonal terms; we use the information criterion from Akaike [1987] without the small-sample correction due to the large number of observations [Macdonald et al., 2018, Table 6.1]. There is little difference in the order of adding the age term in equation (4) or the shape term in equations (5) and (6) — the amplitude of age-related seasonal variation is more significant than the shape.

Table 7: Development of AIC from adding seasonal terms. Source: own calculations using datasets described in Table 1.

Model	CHL	FRA	UK2
No seasonal terms	1,754,993	332,804	480,181
+cosine terms, ζ & τ	-985	-135	-596
+age term, ξ	-78	-46	-27
+shape term, ψ	-52	-10	-11
	1,753,878	332,613	479,547

Figure 9 shows how the residuals by month improve considerably from adding all the seasonal terms. However, Figure 9 also suggests that there is still work to do — the residual for the first twelfth of the year is still strongly positive. In the case of the UK2 data set, this residual is larger than can be explained by random variation, suggesting that the model still does not fully capture the spiky nature of the mortality peak.

Figure 9: Deviance residuals from model fits, showing the better fit from including seasonal terms, albeit with room for further improvement. Source: model fits from first and fourth rows in Table 7.



D An alternative approach to seasonal shape

There are many alternative approaches to equation (6) for modelling the shape of seasonal variation. One equally successful approach we explored was a cubic function centred around the peak that blended into a quadratic function six months later for summer mortality; we used $s((y-\tau) \mod 1)$, with $s(t)$ defined in equation (10).

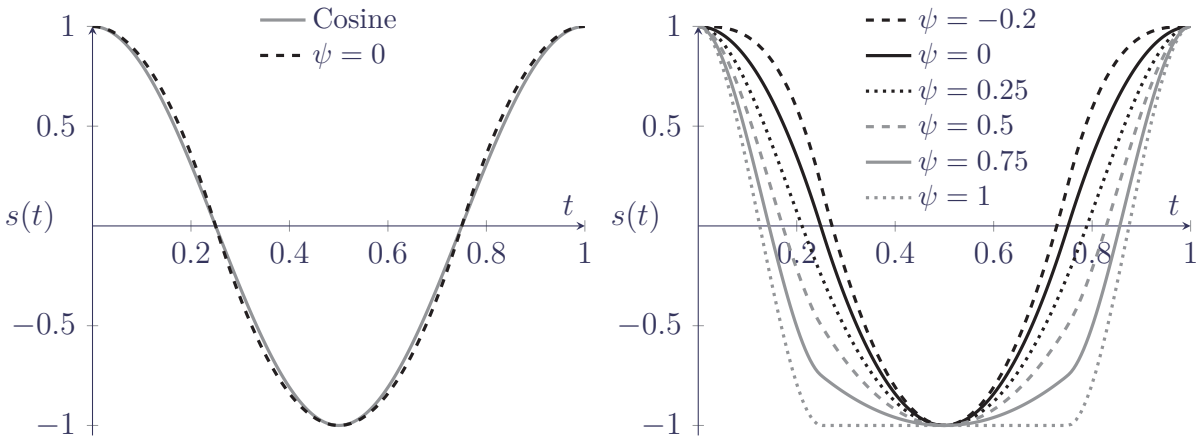
As in equation (6), we desire to have a single parameter, ψ, representing the shape of the seasonal pattern within the year, and which we can estimate from the data. If we use the following coefficients for equation (6):

$$s(t) = \begin{cases} t \in [0, 0.25) & : a + bt + ct^2 + dt^3 \\ t \in [0.25, 0.5] & : e + ft + gt^2 \\ t \in (0.5, 1] & : s(1-t) \end{cases} \tag{10}$$

$$a = 1 \quad b = 0 \quad c = -16(1+5\psi) \quad d = 256\psi \quad e = 3 - 4\psi \quad f = 16(\psi - 1) \quad g = 16(1-\psi) \tag{11}$$

then we have a single winter maximum with $s(0) = 1$ if $\psi > -0.2$ and we have a single summer minimum with $s(0.5) = -1$ if $\psi < 1$. If $\psi \in (-0.2, 1)$ then $s''(0) < 0$ and $s''(0.5) > 0$, and $s(t)$ is C^1 continuous at both $t = 0.25$ and $t = 0.75$. The left panel of Figure 10 shows that when $\psi = 0$ then equation (5) is very similar to equation (3). If ψ should prove to be statistically insignificant, then one could revert to equation (3) or (4). The right panel of Figure 10 shows the range of shapes produced by different values of ψ.

Figure 10: Shape of seasonal effect, $s(t)$, in equation (10). Left: similarity to cosine approach in equation (3) when $\psi = 0$. Right: shape of seasonal effect with different values of ψ.



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).

LONGEVITAS™