


A multi-parameter-level model for simulating future mortality scenarios with COVID-alike effects

Rui Zhou^{1*}  and Johnny Siu-Hang Li²

¹Department of Economics, The University of Melbourne, Melbourne, VIC 3010, Australia; and ²Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada

*Corresponding author. E-mail: rui.zhou@unimelb.edu.au

(Received 26 August 2021; revised 14 February 2022; accepted 25 March 2022; first published online 24 May 2022)

Abstract

There has been a growing interest among pension plan sponsors in envisioning how the mortality experience of their active and deferred members may turn out to be if a pandemic similar to the COVID-19 occurs in the future. To address their needs, we propose in this paper a stochastic model for simulating future mortality scenarios with COVID-alike effects. The proposed model encompasses three parameter levels. The first level includes parameters that capture the long-term pattern of mortality, whereas the second level contains parameters that gauge the excess age-specific mortality due to COVID-19. Parameters in the first and second levels are estimated using penalised quasi-likelihood maximisation method which was proposed for generalised linear mixed models. Finally, the third level includes parameters that draw on expert opinions concerning, for example, how likely a COVID-alike pandemic will occur in the future. We illustrate our proposed model with data from the United States and a range of expert opinions.

Keywords: COVID-19; Mortality scenario generators; Stochastic mortality models; Expert opinions

1. Introduction

The COVID-19 pandemic can be seen as an intervention to long-term mortality dynamics, introducing an additional piece of uncertainty surrounding future mortality experiences of pension portfolios. Traditional stochastic mortality models, which are estimated to (typically a few decades of) historical mortality data, clearly cannot incorporate the uncertainty induced by the pandemic and are thus unable to fulfil the needs of practitioners who wish to envision how future mortality may be affected by the COVID-19 pandemic and perhaps a similar pandemic that may occur in the future. In our view, there is a need to develop a broader framework of stochastic mortality models, which learn from not only historical mortality data but also inputs from other sources.

During the outbreak of COVID-19, information concerning the mortality arising from the pandemic is limited. Death tolls predicted by medical and biological scientists with infectious disease models would enrich stochastic mortality models used in actuarial work, especially in the aspect of disentangling COVID-related and non-COVID-related mortality trends. We call such predictions “external forecasts,” which are warranted to be incorporated into the broader modelling framework.

Although there exist advanced infectious disease models, many determinants of the impact of COVID on future mortality improvements remain impossible to model in a purely data-driven manner. Examples include the long-term effect on health, the potential positive impact due to

improved hygiene, policy changes, and advancement in medical sciences including vaccine development. In this aspect, it would be useful to consider expert opinions, which can complement the information that cannot be extracted from historical data. As a fact, expert opinions have been shown to be useful in strengthening models when data are lacking and modifying models to suit practical needs in various disciplines (Murray *et al.* 2009; Kuhnert *et al.* 2010; R.J. Verrall 2005).

The key contributions of this paper are three-fold. First, we introduce a broader framework of stochastic mortality models that takes both external forecasts and expert opinions into account, on top of the information contained in historical data. Second, we estimate the proposed models with a one-stage penalised quasi-likelihood (PQL) approach which properly disentangles the impact of COVID and the fluctuation in long-term mortality improvements. Third, we examine parameter uncertainty from different sources and illustrate how parameter uncertainty may be incorporated into the simulation of future mortality scenarios.

The specific model we propose encompasses three levels of parameters. The first level includes parameters that capture the long-term, COVID-free pattern of mortality. The second-level parameters reflect the impact of COVID on the basis of external forecasts. The third-level parameters draw on expert opinions on the long-term impact of COVID, a determinant of future mortality improvements that is very difficult to model on a statistical basis. The proposed model allows us to generate stochastic mortality scenarios that incorporate the short- and long-term impacts of COVID as well as the possibility of an outbreak of a COVID-alike pandemic in the future. The scenarios help pension plan sponsors better envision how the mortality experience of their members may continue to evolve.

Several researchers (Chen & Cox 2009; Zhou *et al.* 2013; Lin & Cox 2008; Deng *et al.* 2012; Liu & Li 2015) proposed mortality models that incorporate temporary or permanent jumps, which may be associated with events such as pandemics. These models are built to capture the patterns of historical jump effects, which are then extrapolated into the future. However, each pandemic has its own distinct features, so relying on the information from historical mortality jumps alone may not be effective in predicting the impact of a new pandemic. For example, while half of the deaths due to the 1918 Spanish flu were from the 20–40 age group, the majority of the reported COVID deaths were from the 70-and-over age group. Thus, using the impact of the 1918 pandemic as the sole predictor of the impact of COVID can lead to large errors. Societal changes occurred between the two pandemics may have also altered the speed of transmission. For instance, globalisation has accelerated the spread of diseases, but facilitated an efficient global collaboration towards disease mitigation. We therefore believe that it is important to utilise the newest knowledge of the pandemic when evaluating its impact on future mortality improvements.

Previous studies of mortality jumps, including the works of Lin & Cox (2008), Chen & Cox (2009) and Zhou *et al.* (2013), incorporated mortality shocks into the dynamics of long-term mortality improvement. This setup is restrictive, because the mortality shock caused by a pandemic is forced to share the same age pattern with the long-term mortality improvement. Liu & Li (2015) examined the age pattern of mortality jumps using US data and demonstrated that the age patterns of mortality jumps and long-term mortality improvement are significantly different. On the basis of this empirical finding, they extended the Lee-Carter model to include a separate age pattern for mortality jumps and showed that the extension leads to a better fit to historical data.

While we allow the age patterns of COVID-related and non-COVID-related mortality to be different, our model and findings differ from that of Liu & Li (2015) in a number of aspects. First, Liu & Li (2015) studied historical mortality jumps, the impact of which has been fully observed, and estimated jump-related parameters using historical data; however, we deal with a new mortality shock that is still unfolding and partially understood, so that external forecasts and expert opinions are sought to estimate jump-related parameters. Second, the mortality jumps studied in Liu & Li (2015) are attributable to different causes, so their magnitudes and age patterns may vary. In contrast, we focus on the impact of COVID on mortality and evaluate the uncertainty

arising from different external forecasts. Third, Liu & Li (2015) assume that a mortality jump only affects mortality in its occurrence year, and thus, the impact is transitory; however, we allow the pandemic impact to last for multiple years and also consider permanent shifts in mortality trends caused by COVID.

We view the proposed model as a generalised linear mixed model (GLMM) and estimate its first and second levels of parameters by maximising its PQL. The PQL approach is a single-stage estimation method developed for estimating GLMM by Breslow & Clayton (1993). In the literature, the Lee-Carter model and its variants are often estimated using a two-stage approach, in which the estimates of age and period effects are first obtained and then a process for the estimated period effects is calibrated. However, we find that the two-stage approach cannot properly disentangle the impact of COVID and the fluctuation in long-term mortality improvements. The one-stage PQL approach assumes a multivariate normal distribution for the long-term mortality improvements and penalises the deviation of the estimated long-term mortality improvements from the assumed distribution. The penalty term helps the PQL approach produce reasonable estimates for all of the model parameters. So far as we are aware, our work represents the first attempt to apply the PQL approach to stochastic mortality modelling, although Cairns *et al.* (2011), Haberman & Renshaw (2012), and Liu & Li (2015) have considered other single-stage estimation methods in the same context.

With the first and second levels of parameters estimated using the PQL approach, we determine the third level of parameters using expert opinions. We consider a range of expert opinions on the long-term impact of the COVID and illustrate the long-term impact with simulated mortality paths.

We collect external forecasts from three different studies (Ferguson *et al.* 2020; Centres for Disease Control and Prevention 2020; O'Driscoll *et al.* 2021), conducted several months apart from each other. The three external forecasts are notably different. We examine how such differences may affect the estimated mortality model and the resulting simulated mortality scenarios. This examination highlights the importance of utilising the most recent information in an evolving pandemic. In addition, each external forecast bears a certain extent of forecast uncertainty. We illustrate how this piece of uncertainty may be incorporated into the simulation of mortality scenarios.

The remainder of the paper is organised as follows. Section 2 introduces our proposed model. Section 3 describes the data. Section 4 estimates the proposed model with the PQL method. Section 5 examines how the choice of external forecasts may affect the estimated parameters in the first and second levels. Section 6 explains how expert opinions are used to estimate the third level of parameters. Section 7 discusses how the uncertainty entailed in the external forecasts may be incorporated into the simulation of mortality scenarios. Finally, section 8 concludes.

2. The Proposed Model

2.1. Overview

The proposed model comprises three parameter levels. The first-level parameters capture the long-term pattern of mortality and include the age, period, and cohort components that are also seen in many existing stochastic mortality models including the Lee-Carter model (Lee & Carter, 1992) and the Cairns-Blake-Dowd model (Cairns *et al.* 2006). The first-level parameters are informed by historical mortality data, capturing the dynamics of mortality in the absence of a pandemic like COVID. The second-level parameters gauge the short-term age-specific excess mortality caused by COVID and are calibrated using external forecasts of COVID deaths made in infectious disease studies. Since external forecasts utilise the most recent development of the disease, the second-level parameters update the mortality model with the newest information. The third-level parameters draw on expert opinions concerning the long-term impact of

COVID, including experts' views on how the excess mortality due to COVID may reduce over time, whether COVID will induce positive long-term impact on the society, and how likely a COVID-alike pandemic will occur in the future.

The proposed model, developed for use amid a pandemic, aims at updating mortality models with the newest development in infectious disease studies and enabling users to incorporate the short-term and long-term impact of COVID-alike events into their mortality scenarios. It can also be used post-pandemic, when calibrated to the realised excess mortality. In this paper, we illustrate the proposed model under the assumption that the pandemic is still unfolding and there is significant uncertainty associated with the pandemic impact.

2.2. The long-term mortality pattern without the impact of COVID

We decompose the long-term mortality pattern into age and period effects in the same way as the Lee-Carter model. Historical mortality data in the sample age range of $[x_0, x_1]$ and the sample period of $[t_0, t_1]$ are used to estimate the Lee-Carter components. Let $D_{x,t}$, $E_{x,t}$, and $m_{x,t}$ be the number of deaths, the number of exposures-at-risk, and the central death rate at age x in year t , respectively. We assume that $D_{x,t}$ follows a Poisson distribution with a mean of $E_{x,t}m_{x,t}$. The long-term pattern for $\ln m_{x,t}$ is expressed as follows:

$$a_x + b_x k_t$$

where a_x represents the average log mortality at age x , k_t is the period effect driving long-term mortality improvements over time, and b_x measures the sensitivity of the long-term mortality at age x to changes in k_t . To stipulate parameter uniqueness, we impose the following identifiability constraints: $\sum_x b_x = 1$, and $k_{t_0} = 0$. Although other constraints for the period effects may be used, we choose $k_{t_0} = 0$ to facilitate the one-stage PQL estimation method. This constraint is further discussed in section 4.3.

We assume that $\{k_t\}$ follows a random walk of drift:

$$k_t = k_{t-1} + \mu + \epsilon_t$$

where μ is the drift and $\{\epsilon_t\}$ is a sequence of i.i.d. normal random variables with a mean of 0 and a standard deviation of σ . Other processes such as autoregressive integrated moving average (ARIMA) can also be used if necessary.

The first parameter level contains all of the parameters that are related to long-term mortality dynamics, including a_x , b_x , k_t , μ , and σ . It should be noted that the first-level parameter level may be built on other models such as the Cairns-Blake-Dowd (CBD) model. We choose to use the Lee-Carter structure due to its simplicity.

2.3. Excess age-specific mortality due to COVID

Before the end of a pandemic, it is difficult to predict how long the excessive mortality caused by the pandemic will last. In an early study on COVID mortality, Ferguson *et al.* (2020) argued that COVID-related deaths will mostly occur in 2020 with a large peak if no measure is taken to suppress the transmission of the disease, but spread across two years with smaller peaks if suppression strategies are enforced. In this regard, we believe that it is important to allow for both single-year and multi-year impacts of COVID in the modelling work.

The impact of a pandemic on mortality depends heavily on age. Half of the deaths caused by the 1918 flu pandemic occurred among 20- to 40-year-olds, but older people are the most vulnerable to COVID as shown by various studies (Ferguson *et al.*, 2020; O'Driscoll *et al.*, 2021). The dependence on age is taken into account in our modelling work through a set of age-specific parameters that are reserved for the excess mortality due to COVID.

Assume that the pandemic begins in year T and its impact on mortality lasts for n years. We extend the Lee-Carter structure to capture age-specific excess mortality caused by COVID

as follows:

$$\ln m_{x,t} = a_x + b_x k_t + c_{x,t} \pi_t \mathbb{1}_{\{T \leq t \leq T+n\}} \tag{1}$$

where π_t represents the overall severity of the pandemic in year t , and $c_{x,t}$ captures the age pattern of the pandemic impact in year t . The proposed model assumes that COVID increases central death rates by the multiplicative factor $e^{c_{x,t} \pi_t \mathbb{1}_{\{T \leq t \leq T+n\}}}$. The assumption of a multiplicative impact of COVID on mortality is justified by the apparent proportional relationship between COVID mortality and non-COVID mortality at adult ages, as demonstrated by Cairns *et al.* (2020). By letting $c_{x,t}$ depend on t , we allow the age pattern of the impact of COVID on mortality to change throughout the course of the pandemic. If the user expects that the age pattern should remain unchanged over time, then $c_{x,t}$ can be replaced by c_x .

Theoretically, we can use $c_{x,t}$ instead of $c_{x,t} \pi_t$ to represent the age-specific excess mortality caused by COVID. We choose to use $c_{x,t} \pi_t$ since it shares a similar structure with the long-term mortality improvement term $b_x k_t$ and thus facilitates the comparison between the long-term mortality change and the pandemic impact. This multiplicative structure introduces an identification problem, because given a set of estimates $\hat{c}_{x,t}$ and $\hat{\pi}_t$, we can construct another set of estimates $\hat{c}_{x,t}^* = \hat{c}_{x,t} \times u$ and $\hat{\pi}_t^* = \hat{\pi}_t / u$, the product of which is $\hat{c}_{x,t} \hat{\pi}_t$. To ensure parameter uniqueness, we impose three constraint $\sum_x b_x = 1$, $k_{t_0} = 0$, and $\sum_x c_{x,t} = 1$. Since b_x and $c_{x,t}$ are subject to the same constraint, their estimated values are on the same scale. As a result, the estimate of π_t is comparable to that of k_t and provides a clear indication of the pandemic severity.

The second-level parameters include $c_{x,t}$ and π_t . The value of these parameters is determined by considering external forecasts of excess deaths from infectious disease studies. Forecasts from infectious disease studies are often built on both data collected during the current pandemic and relevant past pandemic experiences. By estimating the second-level parameters using the external forecasts, we inform the model with the new development of the pandemic and also implicitly incorporate information learned from past experience.

2.4. Expert opinions on the long-term impact of COVID

The parameters in the third level gauge the long-term impacts of pandemics, which include the following: excess mortality over an extended period, potential positive impact due to behavioural changes and improved healthcare practice, and possible occurrence of pandemics similar to COVID in the future. Adding the third-level parameters, our proposed model becomes

$$\ln m_{x,t} = a_x + b_x k_t + \sum_i \left(c_{x,t}^{(i)} \pi_t^{(i)} \mathbb{1}_{\{T_i \leq t \leq T_i+n_i\}} + d_{x,t}^{(i)} \psi_t^{(i)} \mathbb{1}_{\{t \geq T_i+1\}} \right)$$

where T_i represents the arrival year of the i th pandemic, n_i is the duration of excess mortality due to the i th pandemic, $c_{x,t}^{(i)}$ and $\pi_t^{(i)}$ capture the age pattern and size of the excess mortality due to the i th pandemic in year t , and $d_{x,t}^{(i)}$ and $\psi_t^{(i)}$ are the age pattern and size of the positive impact induced by the i th pandemic. We assume that the positive impact emerges one year after a pandemic occurs. If expert opinions suggest that the positive impact begins m years after, we can simply set $\psi_t^{(i)}$ to 0 for $T_i + 1 \leq t < T_i + m$. To ensure parameter uniqueness, we impose the following constraints: $\sum_x b_x = 1$, $k_{t_0} = 0$, $\sum_x c_{x,t}^{(i)} = 1$, and $\sum_x d_{x,t}^{(i)} = 1$.

External forecasts of the impact of a pandemic typically focus on short-term excess mortality and are performed using infectious disease models. The long-term impact of the pandemic may not be effectively predicted in the same manner, for reasons including but not limited to virus mutations, policy changes, and medical breakthroughs. We therefore believe that it is more appropriate to estimate the third-level parameters using expert opinions instead of external forecasts.

Some believe that COVID will end with the rollout of the vaccine and the establishment of herd immunity, while others may think that the virus will never be completely eradicated but become a seasonal flu. We do not side with any opinion, but offer an approach to incorporate such opinions into the mortality model. To model how the excess mortality tapers off, we may use a power function or exponential function for π_t . If the expert believes that the pandemic will become a seasonal flu over time, we may change $c_{x,t}^{(i)}$ to match the age pattern of excess mortality caused by a flu.

Noymer (2011) reveals that the 1918 influenza pandemic reduced tuberculosis mortality and transmission since many with tuberculosis were killed in 1918. Cairns *et al.* (2020) argue that COVID accelerates the death of the unhealthy and results in a modest increase of life expectancy among the survivors. While these positive impacts seem temporary, a pandemic may also lead to permanent improvements in mortality due to, for example, behavioural changes and improved health care. Medical research may also be accelerated due to additional investments triggered by the pandemic. We may let $\psi_t^{(i)}$ take a functional form to capture the variation of the positive impact of the i th pandemic over time. The function can be chosen on the basis of expert opinions.

We assume that the arrival of pandemics follows a Poisson process. Equivalently speaking, the interarrival time, i.e., the time between successive pandemics, is exponentially distributed. Assuming that pandemics occur once per λ years on average, the mean interarrival time is λ years and the exponential distribution has a parameter of $1/\lambda$. The distribution of the interarrival time can be written as follows:

$$T_i - T_{i-1} \sim \text{Exponential}(1/\lambda)$$

The value of λ is not easy to determine as pandemics are of a low-frequency nature. We therefore choose λ on the basis of expert opinions. Some scientists including Ross *et al.* (2015) believe that the potential for diseases to spread and the risk of pandemics are increasing due to deepening globalisation and human impact on the environment. To accommodate such opinions, we can adopt a non-homogeneous Poisson process for the arrival of pandemics by letting λ to be time-dependent.

3. Data

Historical US mortality data, including the number of deaths and the number of exposures-at-risk in the period of 1970–2019 and the age range of 20–100, as well as the population size in 2020, are obtained from The Human Mortality Database (2021). Estimated infection fatality ratios (IFR; defined as the number of deaths per infection) are collected from the following three studies.

- The work of Ferguson *et al.* (2020):
This report is produced by the Imperial College COVID-19 Response Team in March 2020. It provides early estimates of COVID IFRs and forecasted number of COVID deaths based on different suppression strategy assumptions.
- The report by the Centres for Disease Control and Prevention (2020):
This report provides five scenarios, which respectively represent different levels of disease transmission, in an effort to advance public health preparedness and planning. The scenarios were first published in March 2020 and are updated every two months since then. We use the set of scenarios published in September 2020.
- The study of O'Driscoll *et al.* (2021):
This study performs a systematic review and meta-analysis of published evidence on COVID-19 mortality until July 2020. Based on the analysis, 15,000 scenarios of IFRs are generated.

The estimates of IFR are derived from best available information at the time of each study. However, due to the evolving nature of COVID, studies performed a few months apart might

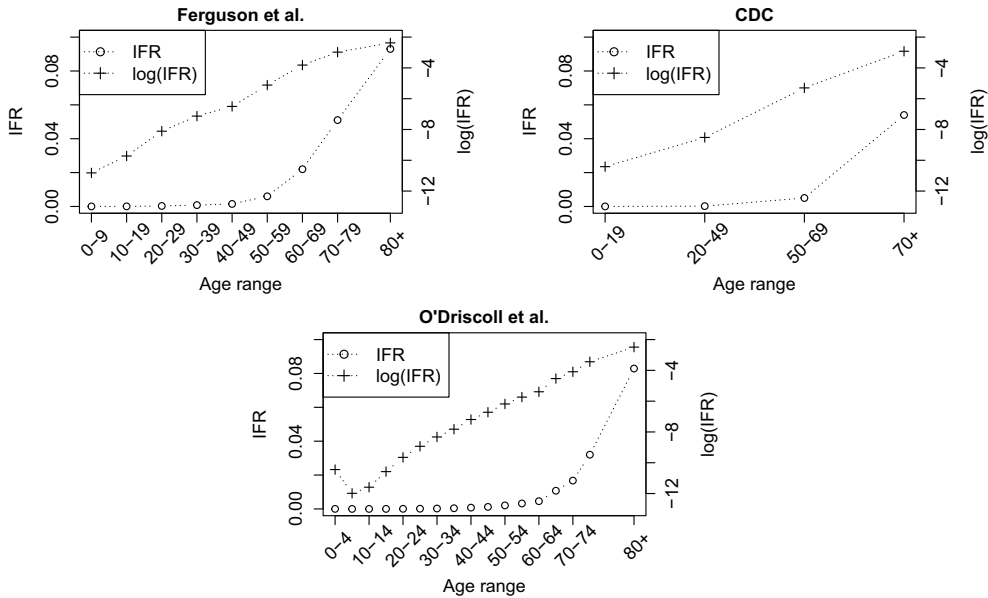


Figure 1. Best estimates of infection fatality ratios (IFRs) produced from three different studies.

result in very different estimates. Biggs & Littlejohn (2020) compare the predicted number of COVID deaths based on the IFR estimates provided by the three studies. The comparison reveals striking differences between the predictions made by Ferguson *et al.* (2020) and those produced by the other two studies, demonstrating how new knowledge may affect the estimate of COVID's impact. In this paper, we consider the three sets of IFR estimates to illustrate the importance of updating mortality models and scenarios with the newest information.

We plot the best estimates of age-specific IFRs from the three studies in their original scale and log scale in Figure 1. All three sets of estimates exhibit an increasing trend with age. The log IFR is approximately a linear function of age, except that for the IFR estimates produced by O'Driscoll *et al.* (2021), the values for the 0–4 age group are higher than those for the 5–9 age group. Since the youngest age group has a very low number of deaths and is generally not covered by insurance or annuities, this exception does not concern us.

We also note that the three studies use different age groupings. To make the three sets of estimates comparable, we interpolate the log scaled IFR using cubic splines. Let us take age groups 0–9 and 10–19 from the Imperial Study as an example. For this study, the estimated log IFRs for the two age groups are 0.002 and 0.006, respectively. We first set the log IFRs of age 5 and age 15 to 0.002 and 0.006, respectively; then for ages between 5 and 15, the estimated log IFRs are obtained by interpolating between 0.002 and 0.006 with a cubic spline. As shown in Figure 2, the interpolated IFRs and log IFRs preserve the increasing pattern in the original values. The estimates from Ferguson *et al.* (2020) are significantly higher than those from the other two studies for the age range of 10–80. The estimates from Centres for Disease Control and Prevention (2020) and O'Driscoll *et al.* (2021) are very close for the age range of 20–80.

We obtain the predicted number of deaths by multiplying the population size of each age at the beginning of 2020, the estimated IFR, and the predicted infection percentage. Ferguson *et al.* (2020) predict an infection percentage of 81% and 510,000 COVID deaths in Great Britain and 2.2 million COVID deaths in the United States in the worst case scenario where no measure is taken against the spread of the virus. The authors also show that the total number of deaths in Great Britain can be reduced to fewer than 10,000 under some suppression strategies. To cover a range

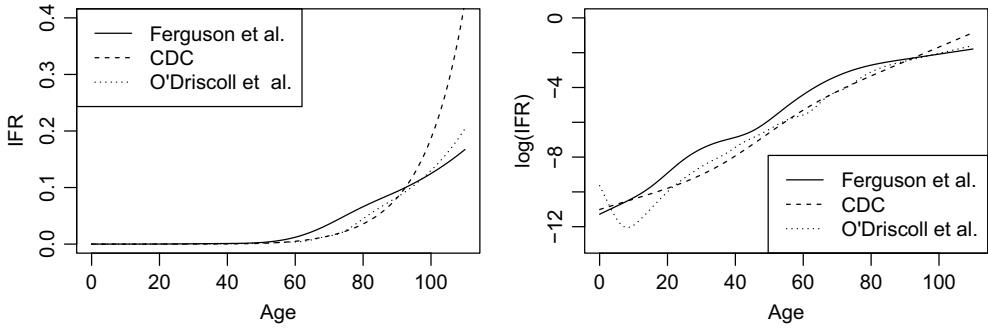


Figure 2. Interpolated infection fatality ratios (IFRs) for individual ages.

of suppression policies, we consider infection percentages ranging from 10% to 80%. We rely on expert opinions to determine which suppression strategy is likely to be used and the resulting infection percentage.

4. Estimating the First Two Levels of Parameters

4.1. Missing data and imputation

In the literature, parameters in the Lee-Carter model are often estimated by a two-stage procedure. In the first stage, parameters representing age and period effects are estimated by methods such as least squares or maximum likelihood. In the second stage, a time-series process is calibrated for the period effects. When applied to our proposed model, the procedure requires the total number of deaths, including both COVID and non-COVID deaths. While we have COVID death estimates from the three studies mentioned in section 3, the total number of non-COVID deaths is still unknown at the time of writing (when our proposed model is estimated). To address this problem, we use multiple imputation to obtain a set of non-COVID death estimates. This method incorporates the uncertainty arising from the missing data by considering a large number of plausible imputed data sets.

Assume that log central death rates in 2020 and onward when COVID is hypothetically absent can be expressed as follows:

$$\ln m_{x,t} = a_x + b_x k_t$$

For simplicity, we assume that the number of deaths due to other causes is not affected by COVID. The following steps are taken to obtain multiple imputed values of $D_{x,t}$ for $t \geq 2020$:

1. Fit the original Lee-Carter model to the 1970–2019 mortality data and obtain estimates of a_x , b_x , and k_t .
2. Estimate a random walk with drift for k_t .
3. Simulate one path for k_t , where $t = 2020, 2021, \dots$, using the estimated random walk with drift.
4. Calculate the corresponding simulated path of $m_{x,t}$, where $t = 2020, 2021, \dots$, using the relation $m_{x,t} = e^{a_x + b_x k_t}$, the simulated path of k_t , and the estimates of a_x and b_x .
5. Calculate simulated values for $q_{x,t}$, where $t = 2020, 2021, \dots$, using the approximation $q_{x,t} \approx 1 - e^{-m_{x,t}}$.
6. Given the simulated value of $q_{x,t}$, simulate the number of non-COVID deaths for age x and year t , $d_{x,t}$, by generating a random number from a binomial distribution with parameters $l_{x,t}$

- and $q_{x,t}$, where $l_{x,t}$ is the population size for age x at the beginning of year t . Then, calculate the population size at the beginning of the following year, using the relation $l_{x,t+1} = l_{x,t} - d_{x,t}$.
7. Repeat the previous step for year $t + 1, t + 2, \dots$
 8. Calculate the imputed total number of deaths for each year t , where $t = 2020, 2021, \dots$, as the sum of the simulated number of non-COVID deaths and the number COVID deaths (obtained from one of the three mentioned external forecasts).
 9. To obtain N sets of imputed death counts, repeat Steps 3–8 N times.

For each set of imputed death counts, we estimate the first and second level of parameters in the mortality model. We will examine the variability of the estimated parameters arising from multiple imputation in section 4.

4.2. A two-stage maximum likelihood estimation method

We first attempt to estimate the model shown in (1) using a two-stage procedure. Assuming that $D_{x,t}$ follows Poisson distribution with a mean of $E_{x,t}m_{x,t}$, the likelihood function can be expressed as

$$\prod_{t,x} \frac{(E_{x,t}m_{x,t})^{D_{x,t}} e^{-E_{x,t}m_{x,t}}}{D_{x,t}!} \propto e^{\sum_{t,x} [D_{x,t} \ln m_{x,t} - E_{x,t}m_{x,t}]}$$

In the first stage, the estimates of $a_x, b_x, k_t, c_{x,t}$, and π_t are obtained by maximising the loglikelihood function

$$ll = \sum_{t,x} [D_{x,t} \ln m_{x,t} - E_{x,t}m_{x,t}] \tag{2}$$

Details of this approach can be found in Brouhns *et al.* (2002). In the second stage, we estimate a random walk with drift for k_t .

We illustrate the estimation using the observed numbers of deaths and exposures-at-risk in 1970–2019 and a set of imputed numbers of deaths in 2020. Only the age range of 20–100 is considered in the estimation. The numbers of COVID deaths used in the imputation are based on the scenario with a 80% infection percentage in 2020 and the best IFR estimates from Ferguson *et al.* (2020). We assume that all COVID deaths occur in 2020. The number of exposures-at-risk in 2020, $E_{x,2020}$, is approximated by $l_{x,2020} - 0.5D_{x,2020}$. The set of imputed numbers of non-COVID deaths and total number of deaths are shown in Figure 3.

Figure 4 presents two sets of parameter estimates obtained from the two-stage estimation procedure. The black lines represent the estimates of a_x, b_x , and k_t in the original Lee-Carter model when the model is fitted the 1970–2019 mortality data. The black dots represent the estimates of a_x, b_x, k_t , and c_x in model (1) when the model is fitted to the 1970–2020 data, in which the age-specific death counts for 2020 are imputed.

Despite the differences in the estimates of a_x, b_x , and k_t obtained from the two sets of data, the 2019 mortality rates implied by the two estimated models are very close to each other for all ages. However, the small difference in the slope of k_t may lead to a large difference in predicted future mortality rates over the long run.

When fitting model (1) to the 1970–2020 data, the estimated series of k_t dips at $t = 2020$. Since k_t represents the mortality level in year t without the influence of the pandemic, we expect the series of k_t to be relatively smooth and follow the downward trend observed in previous periods. The large dip in 2020 is due to the fact that there is no constraint in the optimisation to ensure that k_{2020} would follow the typical pattern of k_t and $c_{x,2020}\pi_{2020}$ would capture the pandemic impact. In particular, the impact of COVID on the 2020 mortality is split between two components, $b_x k_{2020}$ and $c_{x,2020}\pi_{2020}$, both of which capture the interaction of age and period effects, in a way that maximises likelihood.

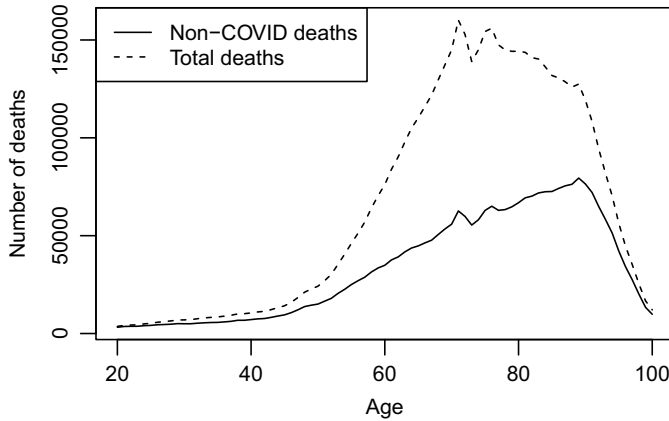


Figure 3. A set of imputed numbers of non-COVID deaths and total deaths in 2020 using a 80% infection percentage and the best IFR estimates from Ferguson *et al.* (2020).

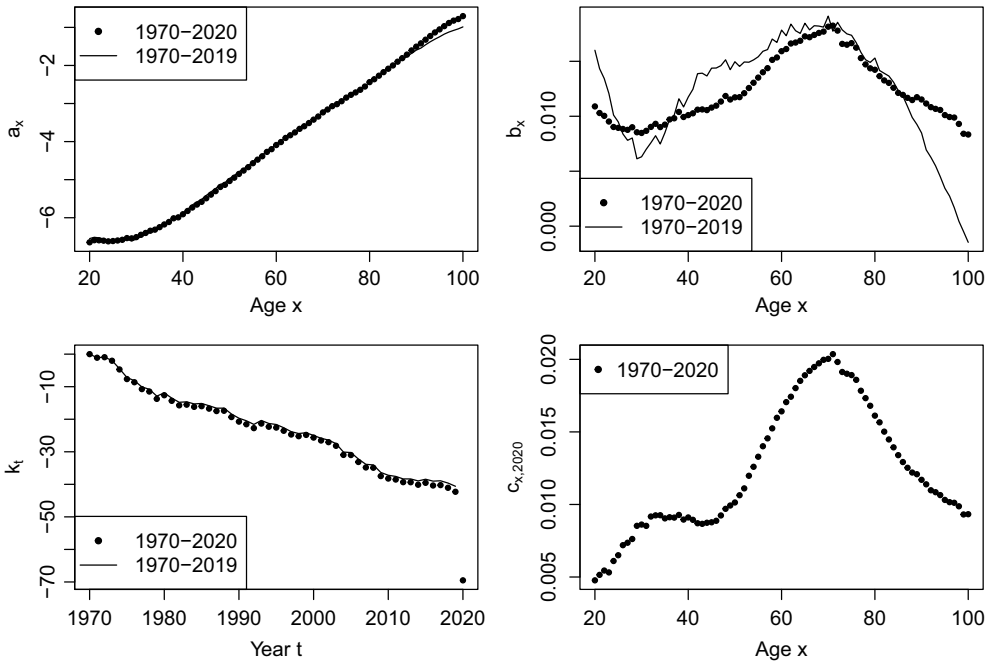


Figure 4. Estimates of a_x , b_x , k_t , and $c_{x,2020}$ obtained from the two-stage estimation procedure.

4.3. Penalised quasi-likelihood estimation

To address the erratic behaviour of k_{2020} , we propose to use a single-stage estimation procedure which simultaneously estimates all first- and second-level parameters, including μ and σ in the random walk with drift for k_t . This procedure is developed to estimate parameters for generalised linear mixed models by Breslow & Clayton (1993). In fact, our model can be written in the form of a generalised linear mixed model (GLMM) with a Poisson distribution for the response variable and a logarithm link function. Although our model differs from a common GLMM in that

we impose parameter constraints on both fixed and random effects, the single-stage estimation procedure can still apply.

Breslow & Clayton (1993) show that the approximation of the marginal quasi-likelihood using Laplace’s method leads to estimating a GLMM based on PQL for the mean parameters. The mean parameters in our model include $a_x, b_x, k_t, \mu, c_{x,t}$, and π_t . Since k_{t_0} is constrained to 0, we do not need to estimate it. Let us denote $[k_{t_0+1}, k_{t_0+2}, \dots, k_{t_1}]'$ by \mathbf{k} . Assuming a random walk with drift for the series of $\{k_t\}$, \mathbf{k} simply follows a multivariate normal distribution with a mean vector $\boldsymbol{\mu}$ and a variance-covariance matrix \mathbf{V} . It is easy to show that

$$\boldsymbol{\mu} = [\mu, 2\mu, \dots, (t_1 - t_0)\mu]', \text{ and}$$

$$V_{ij} = (\min(i, j) - 1)\sigma^2$$

where V_{ij} is the (i, j) -th entry of \mathbf{V} . The joint pdf of the vector \mathbf{k} is denoted by $f(\mathbf{k})$ and can be expressed as

$$f(\mathbf{k}) = 2\pi^{-\frac{t_1-t_0}{2}} |\mathbf{V}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{k}-\boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{k}-\boldsymbol{\mu})}$$

where $|\mathbf{V}|$ is the determinant of the matrix \mathbf{V} . The constraint $k_{t_0} = 0$ not only ensures parameter uniqueness but also initialises the time-series of k_t . Given $k_{t_0} = 0$, it is straightforward to calculate the joint pdf of \mathbf{k} . Although setting $k_{t_1} = 0$ or $\sum_t k_t = 0$ would also lead to unique parameter estimates, the calculation of the joint pdf under such constraints is much more complex.

The integrated (quasi-) likelihood function of our model can be written as

$$\int_{\mathbf{k}} \prod_{t,x} \frac{(E_{x,t} m_{x,t})^{D_{x,t}} e^{-E_{x,t} m_{x,t}}}{D_{x,t}!} f(\mathbf{k}) d\mathbf{k}$$

$$\propto |\mathbf{V}|^{-1/2} \int_{\mathbf{k}} e^{\sum_{x,t} [D_{x,t} \ln m_{x,t} - E_{x,t} m_{x,t}]} e^{-\frac{1}{2}(\mathbf{k}-\boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{k}-\boldsymbol{\mu})} d\mathbf{k}$$

$$= |\mathbf{V}|^{-1/2} \int_{\mathbf{k}} e^{-g(\mathbf{k})} d\mathbf{k}$$

where $-g(\mathbf{k})$ is the PQL function and

$$-g(\mathbf{k}) = \sum_{x,t} (D_{x,t} \ln m_{x,t} - E_{x,t} m_{x,t}) - \frac{1}{2}(\mathbf{k} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{k} - \boldsymbol{\mu}) \tag{3}$$

When a Poisson distribution is used for the response variable, the quasi-likelihood is equivalent to the true likelihood¹. The PQL differs from the loglikelihood functions shown in (2) only by the penalty term $-\frac{1}{2}(\mathbf{k} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{k} - \boldsymbol{\mu})$.

Breslow & Clayton (1993) show that the estimates of the mean parameters in a GLMM can be obtained by maximising the PQL and the obtained estimates approximately maximise the integrated quasi-likelihood. With the mean parameters substituted by their estimated values, the variance parameter σ can be obtained by maximising the approximate profile quasi-likelihood function which is expressed as follows:

$$-\frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} \ln |g''(\tilde{\mathbf{k}})| - g(\tilde{\mathbf{k}})$$

¹Assume that the i th observation of a random variable Y is y_i . The expectation and variance of y_i is $E[y_i] = \mu_i$ and $\text{Var}(y_i) = \phi a_i v(\mu_i)$, where $v(\cdot)$ is a specified variance function, a_i is a known constant, and ϕ is a dispersion parameter that may or may not be known. Given that Y follows a Poisson distribution, we have $\phi = a_i = 1$ and $v(\mu_i) = \mu_i$. The quasi-loglikelihood for the i -th observation is defined as

$$Q(\mu_i | y_i) = \int_{y_i}^{\mu_i} \frac{y_i - u}{a_i v(u)} du = \int_{y_i}^{\mu_i} \frac{y_i - u}{u} du = y_i \ln \mu_i - \mu_i - y_i \ln y_i + y_i$$

This is equivalent to the true loglikelihood for y_i . Given that y_i follows a Poisson distribution with a probability mass function $\mu_i^{y_i} e^{-\mu_i} / y_i!$, the true loglikelihood is $y_i \ln \mu_i - \mu_i - \ln(y_i!)$. The quasi and true loglikelihood only differ by a constant.

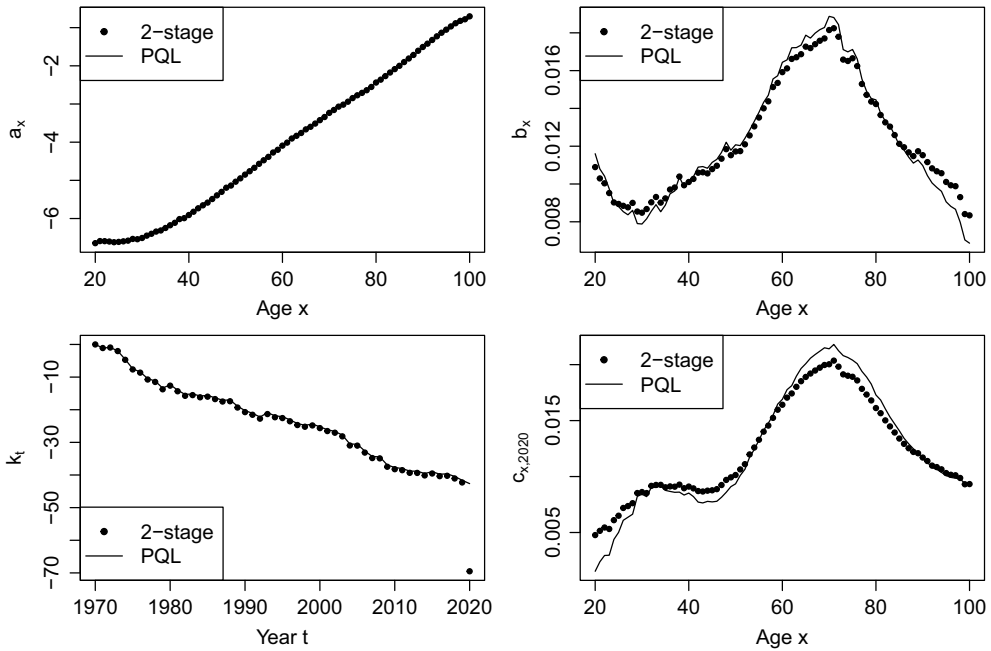


Figure 5. Comparison between the estimates of a_x , b_x , k_t , and $c_{x,2020}$ obtained from the two-stage estimation method and the single-stage PQL approach.

where \tilde{k} is the estimated value of k . The following iterative procedure is taken to obtain the final estimates:

1. Set initial values for a_x , b_x , k_t , $c_{x,t}$, π_t , μ , and σ .
2. Estimate a_x , b_x , k_t , $c_{x,t}$, π_t and μ by maximising the PQL. The optimum is obtained using an iterative Newton’s method with a tolerance level of 0.0001.
3. Take a Newton step towards a new value of σ with the objective of maximising the approximate profile quasi-likelihood function.
4. Repeat steps 2 and 3 until the change of the profile likelihood is smaller than a tolerance value, which we set to 0.0001.

The iterative procedure is implemented with the authors’ own codes. Although the MASS (Venables & Ripley 2002) package can perform the PQL estimation for a GLMM, it cannot handle our parameter constraints and thus is not suitable for estimating the proposed model. The parameter estimates obtained from the two-stage procedure and the PQL method are compared in Figure 5. The estimated series of k_t exhibits no visible dip in year 2020 when the PQL approach is used, suggesting that the one-stage estimation approach is capable of disentangling the impact of COVID and long-term mortality improvements. The two procedures lead to similar estimates for a_x , b_x , and $c_{x,2020}$. However, the estimated value of π_t obtained from the one-stage procedure is 45.62, while that obtained from the two-stage procedure is 60.74. Since π_t represents the severity of the pandemic, the two-stage procedure seems to overestimate the impact of the pandemic and simultaneously underestimate the value of k_t in 2020.

We also observe from Figure 5 that the estimates of $c_{x,2020}$ do not always increase with age. Our model assumes a multiplicative impact of COVID on mortality. The multiplicative factor $e^{c_{x,t}\pi_t}$ is the ratio of the mortality rate with the impact of COVID to that without. Although both rates increase with age, the ratio of the two rates does not necessarily increase with age.

The PQL estimation procedure can also be used to calibrate the realised impact of COVID when the total number of deaths are fully observed. Its advantage of disentangling the impact of COVID and long-term mortality improvements remains in this situation.

The PQL method is similar to the penalised likelihood estimation approach proposed by Delwarde *et al.* (2007). Both approaches add a penalty term to the likelihood function shown in (2). Delwarde *et al.* (2007) achieve smoothed age-specific patterns b_x at the expense of likelihood. The approach of Delwarde *et al.* (2007) has been applied to smooth other parameters, including a_x and k_t in mortality models by Li *et al.* (2020). Although we may use penalised likelihood estimation to smooth the estimates of k_t and avoid the dip at $t = 2020$, the volatility of the estimated k_t would be significantly reduced in this way. Consequently, future mortality scenarios simulated from the estimated model would understate the true volatility in mortality improvements as well as the longevity risk inherited in pension and annuity portfolios. In contrast, the PQL approach considers how closely the vector \mathbf{k} follows a multivariate Gaussian distribution, so it preserves the variability in the estimated k_t . The PQL approach is therefore more suitable for estimating our model, which is developed for users to gauge the range of possible mortality outcomes in the future.

We note that penalising the deviation from a multivariate normal distribution does not guarantee that the estimated k_t follows a multivariate normal distribution. We may consider a stiffer penalty, $-\frac{1}{2}\rho(\mathbf{k} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{k} - \boldsymbol{\mu})$, where ρ is a tuning parameter and $\rho \geq 1$. The estimated k_t will be more likely to follow a multivariate normal distribution as ρ increases. However, this comes at the cost of the goodness-of-fit and the theoretical foundation of the PQL approach. How to choose the tuning parameter is also a challenging question. Since $\rho = 1$ works well for our dataset, we do not consider other values of ρ .

McCulloch & Neuhaus (2011) examine the estimation robustness to the assumed distribution for a random effect when using maximum likelihood methods for fitting a GLMM. They find that the empirical probability density of the estimated random effects can be quite sensitive to the assumed distribution. Since maximising PQL approximates maximising the true likelihood, we expect that the assumed distribution for k_t will also have a significant impact on the distribution of the estimated k_t . Therefore, it is important to assume a suitable distribution for k_t .

Normal distribution is often assumed for the period effect k_t in existing literature. To justify the normal assumption, we first fit a Lee-Carter model to the 1970–2019 mortality data with the two-stage estimation approach and then perform the Shapiro–Wilk normality test to the difference of the estimated k_t . Since the estimation of k_t with the two-stage approach does not require any distribution assumption for k_t and no pandemic occurred during 1970–2019, the distribution of the estimated k_t should be close to the true distribution of k_t . The Shapiro–Wilk test returns a p -value of 0.44 which indicates that we cannot reject normality. Therefore, the normal assumption for k_t seems reasonable.

Alternative to the PQL approach, we may consider a Bayesian approach and view the proposed model as a Bayesian hierarchical model. To implement the Bayesian estimation for our model, we should assume a normal prior for k_t with parameters μ and σ treated as hyperparameters. The hyperparameters are assumed to be random variables with their own prior distributions. Prior distributions are also required for a_x , b_x , and pandemic-related parameters. Inferences are then drawn from the derived posterior distributions. Since a posterior distribution is influenced by the assumed prior distribution to some extent, the normal prior plays a similar role with the penalty term in PQL in separating the changes in the long-term mortality improvement from the pandemic shock. A prior sensitivity analysis is necessary to examine the impact of the normal prior on the posterior. It would be interesting in the future research to carry out a Bayesian estimation of our model and compare the estimates obtained from the two approaches.

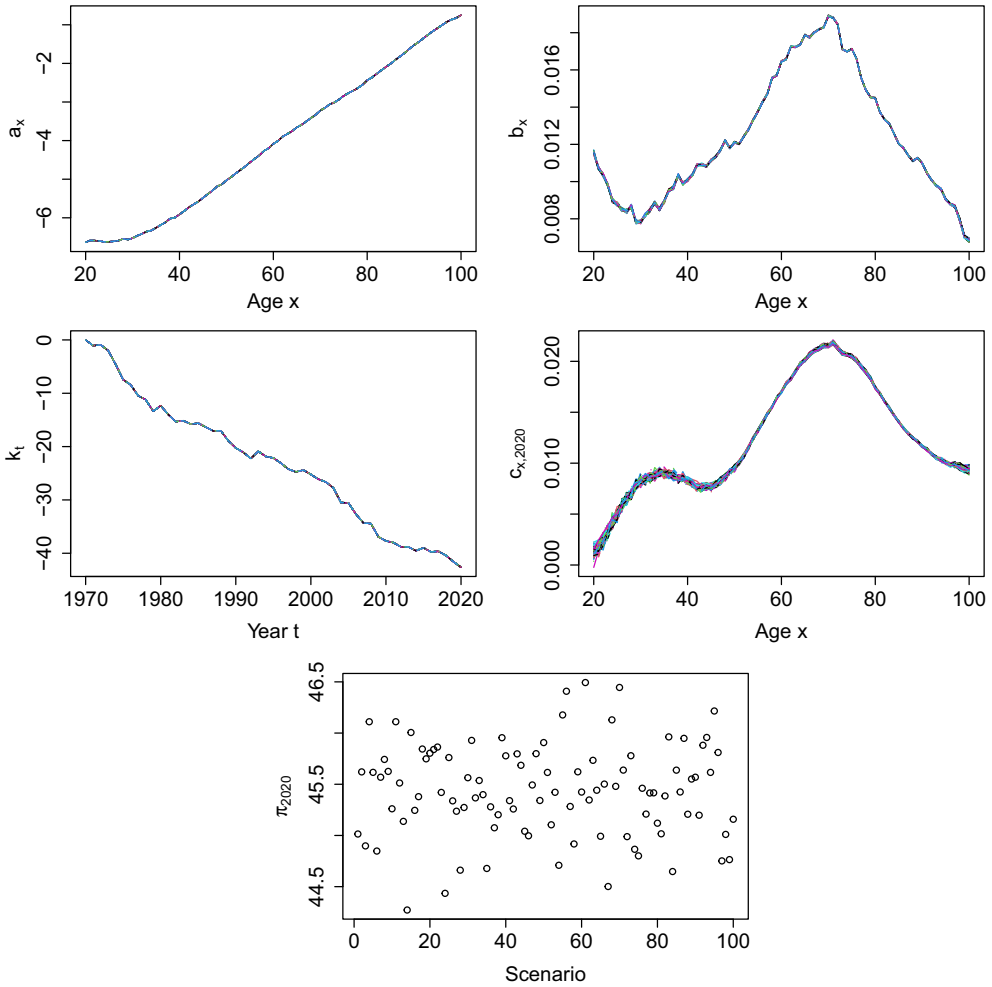


Figure 6. Estimates of a_x , b_x , k_t , $c_{x,2020}$, and π_{2020} obtained using 100 sets of imputed non-COVID death counts in 2020.

4.4. Combining estimates from multiple imputations

Using the multiple imputation procedure described earlier, we obtain 100 sets of imputed age-specific death counts in 2020. The variability of the imputed values arises from the randomness in k_{2020} and $D_{x,2020}$. To examine how imputation affects the estimated parameters, we repeat the one-stage estimation procedure for each set of imputed values and plot the estimated parameters in Figure 6.

Figure 6 shows that different sets of imputed values result in similar estimates of a_x , b_x , and k_t . Therefore, the variability in the imputed values has a minimal impact on the estimates of the first-level parameters. However, the values of $c_{x,2020}$ for young ages and very high ages somewhat vary across different sets of imputed values, because at these ages COVID-related deaths are small so that the impact of the COVID on mortality is more difficult to quantify.

The uncertainty arising from the imputation can be viewed as parameter risk. We can incorporate this piece of uncertainty into simulated future mortality paths by taking the following steps:

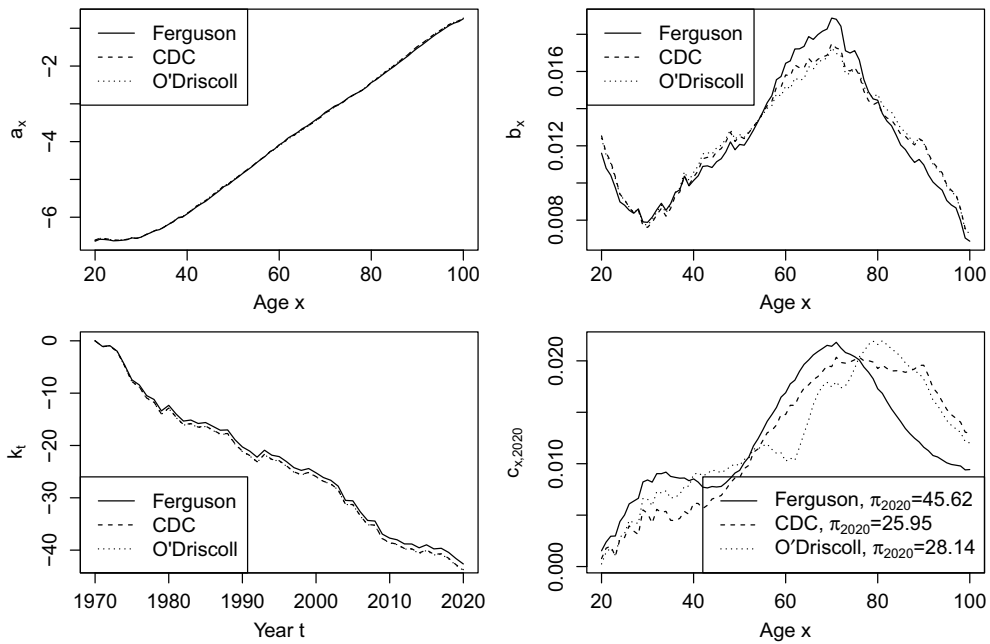


Figure 7. Estimates of a_x , b_x , k_t , $c_{x,2020}$, and π_{2020} , obtained using the IFR estimates provided by Ferguson *et al.* (2020), Centres for Disease Control and Prevention (2020), and O'Driscoll *et al.* (2021).

1. Obtain a set of imputed non-COVID death counts.
2. Estimate the mortality model based on the set of imputed values.
3. Simulate 100 mortality paths using the estimated parameters in Step 2.
4. Repeat steps 1–3 for 100 sets of imputed values.

We find that the randomness in simulated k_t dominates the uncertainty in the simulated mortality paths. Incorporating the uncertainty from the imputed value in the simulation does not lead to any visible change in the resulting prediction intervals.

5. Choice of External Forecasts

We observe significant differences in the estimated IFRs from the three studies in Figure 2. In this section, we examine how the estimates of $c_{x,t}$ and π_t are affected by the source of IFRs. We use the same set of imputed non-COVID deaths in 2020 but three different sets of estimated COVID deaths which are derived from an assumed infection percentage of 80% and the best IFR estimates from the three studies. Figure 7 shows, for each source of IFRs, the estimated values of the first and second levels of parameters (including the estimated value of π_{2020} that is displayed in the legend of the bottom-right panel). Although the three sets of IFRs result in small differences in the estimated first-level parameters, the fitted non-COVID mortality in 2020, expressed as $e^{a_x+b_x k_{2020}}$, resulting from the three sets of IFRs are very close to one another. Therefore, the difference in the first-level parameters does not affect the comparability of the second-level parameters obtained using different IFR estimates.

The severity of the impact of COVID, represented by π_{2020} , is the highest when model is calibrated using the IFR estimates from Ferguson *et al.* (2020). On the other hand, the values of π_{2020} derived from the IFR estimates provided by Centres for Disease Control and Prevention (2020)

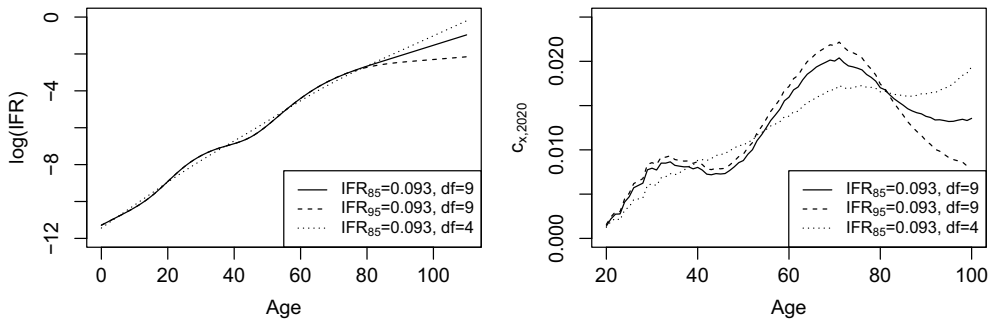


Figure 8. Log IFRs and estimates of $c_{x,2020}$ when three different cubic spline specifications are used.

and O’Driscoll *et al.* (2021) are similar. These estimation results are consistent with the observation of the interpolated IFRs shown in Figure 2. For all three sets of IFR estimates, the resulting estimates of $c_{x,2020}$ generally increase with age first and decrease after a certain age. The three sets of estimated $c_{x,2020}$ differ the most at high ages.

We find that the estimates of $c_{x,2020}$ at high ages are sensitive to the way in which we obtain the cubic spline. For example, the best IFR estimates of Ferguson *et al.* (2020) are 2.5% and 9.3% for the age ranges of 70–79 and 80+, respectively. When implementing the cubic spline method, the IFRs at individual ages are required. For the age range of 70–79, it is reasonable to assign the 2.5% IFR to age 74.5, which is the mid-point of the age range.

However, for the 80+ age range, it is not so clear as to which age should be used as the last knot and assigned the 9.3% IFR. As shown in Figure 8, switching the last knot from 85 to 95 would lead to significant changes in the smoothed log IFRs and the estimates of $c_{x,2020}$ for ages 80 and above. In addition, when the degrees of freedom of the cubic spline are reduced from 9 to 4, the values of the smoothed log IFR would change slightly but the estimates of $c_{x,2020}$ would exhibit significantly less curvature. If possible, we suggest using IFRs for finer age groups to reduce the needs for interpolation and resulting errors.

The estimates of $c_{x,2020}$ are more jagged in the age range of 20–40. This observation is consistent with the larger fluctuation in $c_{x,2020}$ at younger ages, as shown in Figure 6. We believe that the jaggedness is caused by the small number of COVID deaths at these ages, so that it is more difficult to distinguish between the impact of COVID and variation in long-term mortality improvements.

It is evident that the choice of external forecasts matters. A single source of external forecasts may under- or overestimate the impact of COVID and hence future mortality. Combining the information from multiple sets of external forecasts may alleviate the problem. A similar method, which makes use of predictions from 32 models, has been used by the CDC during the pandemic to create short-term projections of COVID cases and deaths. We will further discuss the incorporation of multiple external forecasts into the simulation of future mortality paths in section 7.

6. Using Expert Opinions to Determine the Third-Level Parameters

6.1. Infection percentage

The results presented in the previous sections are obtained using an assumed infection percentage of 80%. The infection percentage may change significantly, if different suppression strategies are adopted. As policy changes cannot be statistically modelled, we consider a number of possible infection percentages and examine the resulting parameter estimates for each. The specific choice of the assumed infection percentage is a subjective judgement, which may be informed by expert opinions.

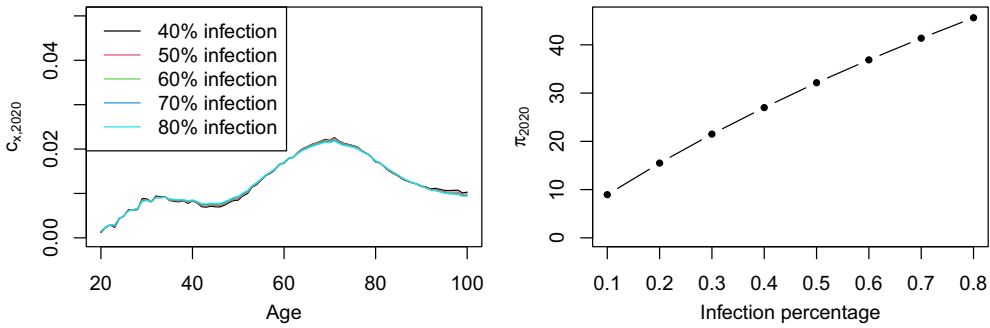


Figure 9. Estimates of $c_{x,2020}$ and π_{2020} obtained using the best IFR estimates from Ferguson *et al.* (2020) and various assumed infection percentages.

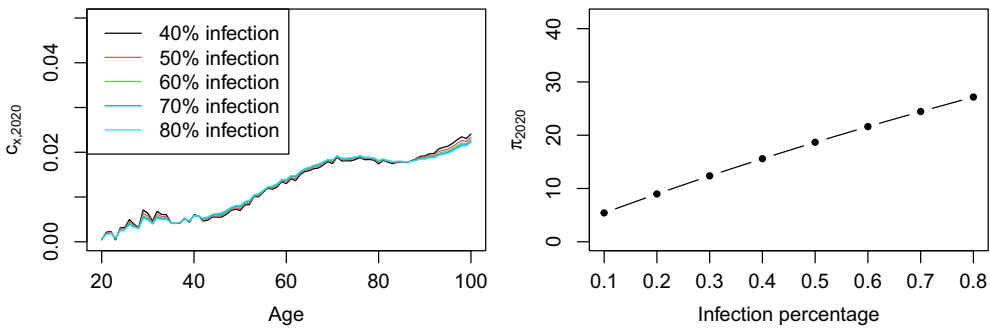


Figure 10. Estimates of $c_{x,2020}$ and π_{2020} obtained using the best IFR estimates from the Centres for Disease Control and Prevention (2020) and various infection percentages.

The left panel of Figure 9 presents the estimates of $c_{x,2020}$ and π_{2020} obtained using the best IFR estimates from Ferguson *et al.* (2020) and various assumed infection percentages. We observe that the estimates of $c_{x,2020}$ only change slightly as the assumed infection percentage increases. The right panel of Figure 9 plots the estimate of π_{2020} against the assumed infection percentage. The estimate of π_{2020} increases with the assumed infection percentage in a close to linear fashion. Similar observations can be made when considering IFR estimates from Centres for Disease Control and Prevention (2020) and O’Driscoll *et al.* (2021), as shown in Figures 10 and 11. These observations can be accounted for as follows. If the infection percentage increases while the age-specific IFRs remain unchanged, then the projected overall severity of the pandemic would increase but the distribution of the impact across ages would not be affected. As a consequence, parameter π_{2020} , which measures the overall fatality level and infectiousness of COVID in 2020, increases with the assumed infection percentage, but the pattern of $c_{x,2020}$, which can be viewed as the standardised age-specific pattern of the impact of COVID, remains largely unchanged.

Figures 9, 10, and 11 also suggest a simple method to adjust the estimated model when the infection percentage changes while IFRs remain the same. Let us suppose that expert opinions suggest that the infection percentage would be reduced to 25% due to new restrictions. The new value of π_{2020} can be approximated by a linear interpolation between the values of π_{2020} corresponding to infection percentages of 20% and 30%. The approximation yields 18.50 as an estimate for π_{2020} , while re-estimating the model completely produces an estimate of 18.58. The approximation method appears to be sufficiently accurate and may be used to save the computational burden of a complete re-estimation.

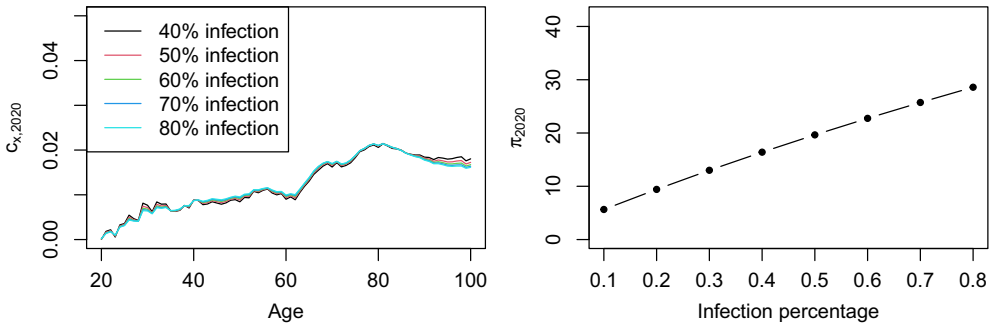


Figure 11. Estimates of $c_{x,2020}$ and π_{2020} obtained using the best IFR estimates from O’Driscoll *et al.* (2021) and various infection percentages.

Although we assume in this paper that the infection percentage is age invariant, we are aware that researchers such as Davies *et al.* (2020) have demonstrated age-dependent effects in the transmission and control of COVID. Should age-specific infection percentages be available and used as an input of our proposed model, the dependence between age and infection percentage would be reflected in the series of $c_{x,t}$.

6.2. Long-term impact of COVID

It is very challenging to predict the long-term impact of COVID due to the uncertainty concerning vaccination, virus mutation, and government policies. In what follows, we present three possible scenarios of the long-term impact of COVID, taking into account of a range of assumed expert opinions. In these scenarios, we assume that IFRs do not change over time but the infection rate decreases gradually. Each scenario is illustrated by a simulated path of $m_{70,t}$ for t beyond 2020. The model used in the simulation is estimated based on the best IFR estimates from Ferguson *et al.* (2020) and a 40% infection rate in 2020.

Scenario 1: Excess mortality tapers off and mortality dynamics revert back to the long-term COVID-free trend

Let F_t be the infection rate in year t . We use a function of time to describe the reduction in infection rate over time. A possible choice is a power function expressed as $F_t = F_{2020}\delta^{t-2020}$, where $0 < \delta < 1$. The value of π_t for $t > 2020$ can be approximated by the linear interpolation described in section 6.1. When such a power function is used, F_t approaches to 0 as t increases, so that the impact of COVID will not completely disappear but will reduce to a negligible level.

In the top left panel of Figure 12, we show simulated paths of $m_{70,t}$ beyond $t = 2020$, for different values of δ in the power function for F_t . For comparison, we also plot a simulated path that is not subject to any COVID impact. As the value of δ increases, it takes longer for the mortality dynamics to revert to the long-term COVID-free trend. Using $\delta = 0.1$, reversion is almost complete in 4 years; but with δ set to 0.9, the impact of COVID is still visible in 30 years.

Scenario 2: Excess mortality reduces for several years and then remains constant

In this scenario, the impact of COVID is long-lasting, but will become much lighter compared to the 2020 level due to, for example, vaccination. This scenario echoes the view of Oliver (2020), who asserts that COVID may have long-term health effects that survivors may suffer from. Assuming that the excess mortality caused by COVID reduces for n years and then remains level, a suitable function for modelling the residual effect of COVID is $F_t = F_{2020}\delta^{\min(t-2020,n)}$.

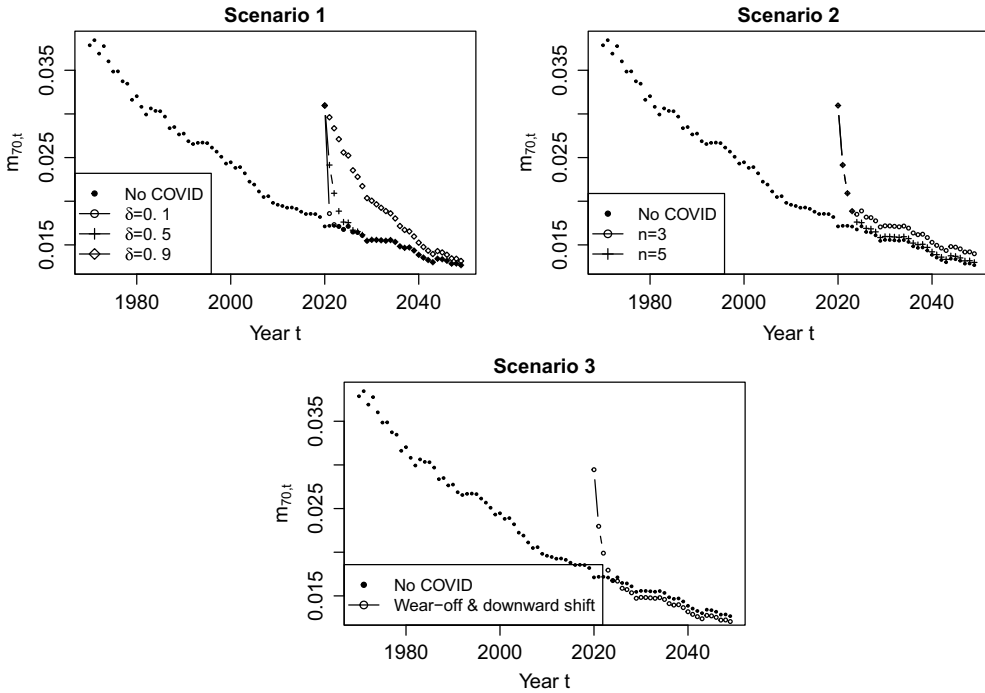


Figure 12. Simulated paths of $m_{70,t}$ for 2021 and beyond in the three scenarios of long-term COVID impact under consideration.

In the top right panel of Figure 12, we plot a simulated path of $m_{70,t}$ with $\delta = 0.5$ and n set to various values. When n is set to 3, excess mortality ceases to reduce in year 3 and significant excess mortality remains afterwards. When n is raised to 5, the reduction in excess mortality lasts longer and the residual excess mortality after the cessation of reduction in year 6 is smaller.

Scenario 3: Excess mortality gradually reduces and a permanent downward shift in mortality occurs due to positive behavioural changes and improved healthcare practice

This scenario combines the wear-off of excess mortality and a permanent downward shift in mortality. The downward shift is represented by $d_{x,t}^{(i)} \psi_t^{(i)} \mathbb{1}_{\{t \geq T_i+1\}}$ in the mortality model. The positive impact may be driven by behavioural changes and changes in health care system (Edwards, 2020). The positive impact induced by COVID is very difficult to be modelled statistically, so expert opinions are sought to determine the relevant parameters.

We illustrate this scenario by assuming that expert opinions indicate a positive impact of COVID will begin from 2021 and will be uniform across all ages. Accordingly, we set $d_{x,t}^{(i)} \psi_t^{(i)} = -0.05$ and $\delta = 0.5$. A simulated path of $m_{70,t}$ obtained from this setting is shown in the bottom panel of Figure 12. We observe that the death rate reduces continuously and its trajectory becomes lower than the long-term COVID-free pattern in 4 years. The positive impact is permanent, affecting all death rates starting from 2021.

6.3. COVID-alike pandemics in the future

As pandemics are very rare events, a reliable statistical estimation of the expected inter-arrival time of pandemics is not possible. To illustrate the impact of possible COVID-alike pandemics in the

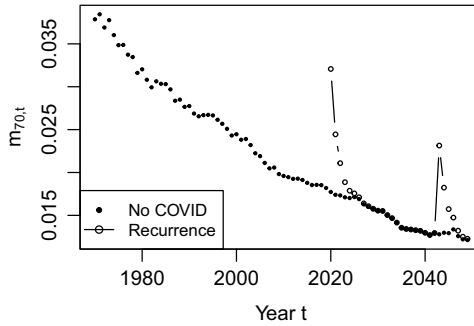


Figure 13. A simulated path of $m_{70,t}$ beyond 2021, under the assumption that COVID-like pandemics will occur at a rate of $\lambda = 1/100$.

future, let us suppose that expert opinions indicate that pandemics like COVID are 1-in-100-years events. Accordingly, we set λ to 1/100.

Figure 13 presents a simulated path of $m_{70,t}$, which is perturbed by not only COVID but also a COVID-like pandemic that occurs in 2029. In the simulation, the wear-off parameter δ is set to 0.5. Although the same values of $c_{x,t}$ and π_t are used for the 2020 and 2029 pandemics, the size of jump in $m_{70,t}$ is much lower for 2029 than 2020. This outcome is due to the fact that our model implies a multiplicative pandemic effect, as discussed in section 2.4. Specifically, our model implies that a mortality rate that incorporates pandemic impacts is the product of the corresponding mortality rate that is not subject to any pandemic impact and a multiplicative factor of $e^{c_{x,t}\pi_t}$. As the pandemic-free mortality $m_{70,t}$ in 2029 is lower than that in 2020 due to long-term mortality improvements, the absolute value of excess mortality in 2029 is lower than that in 2020.

We further present the fanplot for 10,000 simulated paths of $m_{70,t}$ in Figure 14. The paths shown in the left panel incorporate the possibility that COVID-like pandemics will occur in the future, while those shown in the right panel do not. The median, 80th percentile, and 90th percentile appear to be very similar between the two panels. However, the 95th and 99th percentiles in the left panel are very spiky due to the possible future pandemics. As the frequency of pandemics is very low, only the sample paths corresponding to the worst scenarios are affected.

7. Incorporating Parameter Uncertainty into the Simulation of Future Mortality

In section 4, we observe that the IFR estimates from the three different studies under consideration result in significantly different estimates of $c_{x,2020}$ and π_{2020} . It should be noted that the IFRs provided by the three studies are subject to estimation uncertainty, an issue that we now investigate.

In the report of the Centres for Disease Control and Prevention (2020), lower and higher estimates of IFRs are provided. We use them to approximate 95% confidence intervals for the IFR estimates of the Centres for Disease Control and Prevention (2020).

O'Driscoll *et al.* (2021) generate 15,000 simulated IFR values for each age range. We approximate confidence intervals for the IFR estimates of O'Driscoll *et al.* (2021) using the 2.5th and 97.5th percentiles of the 15,000 simulated IFRs.

Ferguson *et al.* (2020) provide only best IFR estimates, which are obtained by adjusting the values in the work of Verity *et al.* (2020), but not any measure of uncertainty surrounding the IFR estimates. We approximate 95% confidence intervals for the IFR estimates of Ferguson *et al.* (2020) by making a proportional adjustment to the 95% confidence intervals of the IFRs provided by Verity *et al.* (2020). The proportion is calculated as the ratio of the best IFR estimates of Ferguson *et al.* (2020) to those of Verity *et al.* (2020).

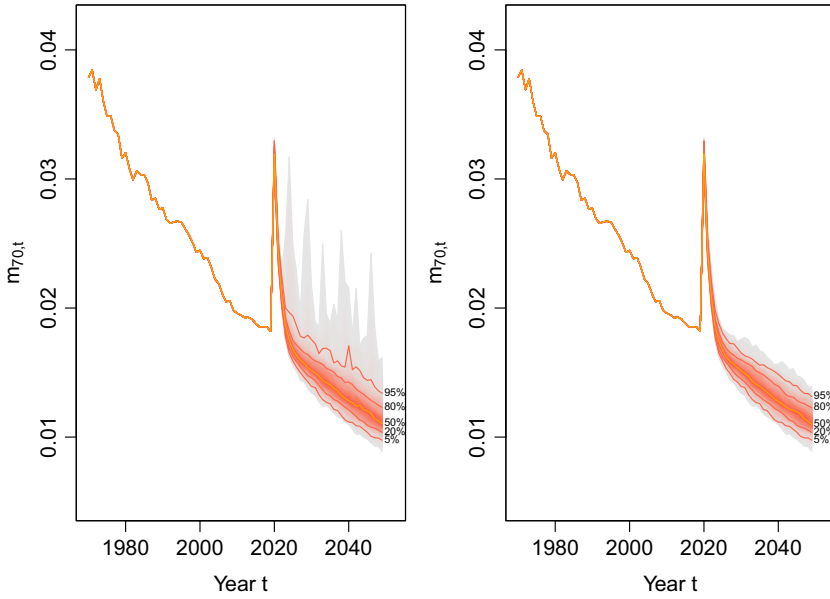


Figure 14. Simulated paths of $m_{70,t}$ under the assumptions that COVID-like pandemics will occur (left panel) and will not occur (right panel) in the future.

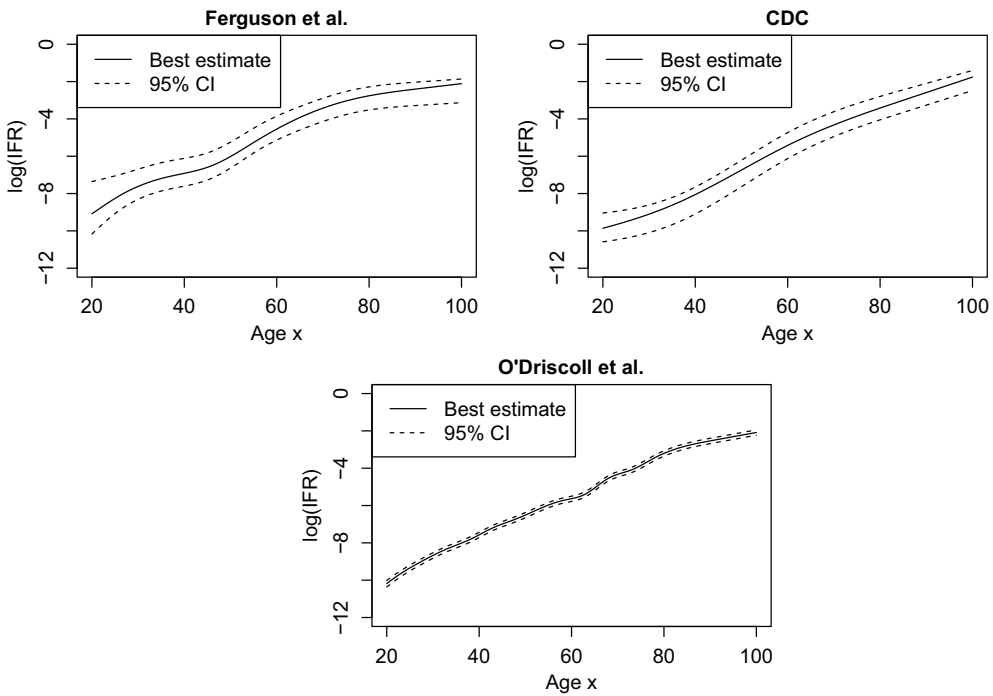


Figure 15. Best estimates and 95% confidence intervals of IFRs from the three studies under consideration.

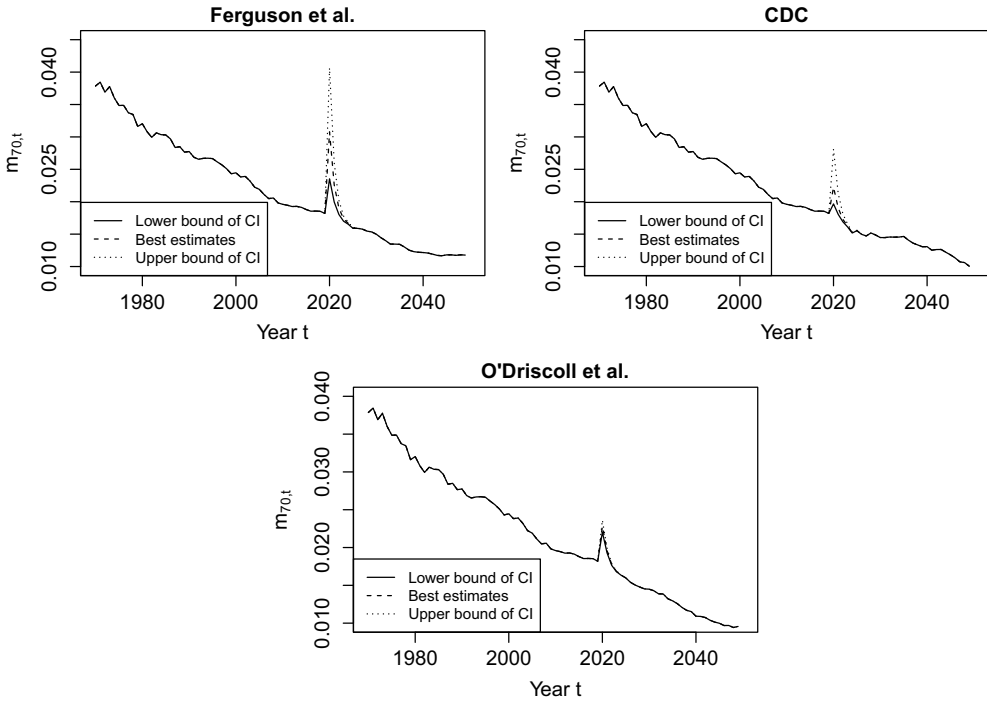


Figure 16. Impact of the uncertainty in the IFR on simulated mortality.

In Figure 15, we plot the best estimates and 95% confidence intervals of IFRs after interpolation from the three studies. The confidence intervals for the IFR estimates of Ferguson *et al.* (2020) and the Centres for Disease Control and Prevention (2020) are significantly wider than that of O’Driscoll *et al.* (2021).

We now examine how the uncertainty surrounding the estimated IFRs may affect simulated long-term mortality. In Figure 16, we compare the simulated paths of $m_{70,t}$ that are obtained using the values of $c_{x,2020}$ and π_{2020} calibrated to IFRs at their best estimate levels, and upper and lower limits in their respective confidence intervals. In the simulation, it is assumed that the excess mortality caused by COVID wears off at a rate of $\delta = 0.5$ and that COVID-alike pandemics will not occur in the future. Also, to focus on the uncertainty surrounding the IFR estimates, we use the same simulated path of k_t for all calculations.

Let us compare the top left panel in Figure 16 and the right panel in Figure 14. The sample paths in both diagrams are obtained using the IFRs of Ferguson *et al.* (2020). In the absence of the uncertainty arising from the IFR estimates, the variation in $m_{x,t}$ comes solely from the randomness of k_t , and as shown in the fanplot in right panel of Figure 14, the confidence interval of $m_{x,t}$ for the first few years after the commencement of COVID is narrow. The result shown in Figure 16 suggests that if the uncertainty arising from the IFR estimates is incorporated, then the uncertainty arising from the IFR estimates would be the primary source of forecast uncertainty for the first few years after COVID; however, as the excess mortality due to COVID wears off, the randomness of k_t would take over as the major source of forecast uncertainty.

Similar arguments can be made for the IFR estimates of the Centres for Disease Control and Prevention (2020). However, as the confidence intervals for the IFR estimates of O’Driscoll *et al.* (2021) are very narrow, we expect that the randomness of k_t would always be the major source of forecast uncertainty when these IFR estimates and their measures of uncertainty are taken as input.

We can also use simulation to incorporate all sources of uncertainty: k_t , possibility of COVID-alike pandemics in the future, imputation, and IFR estimates by taking the following steps:

1. Select a set of simulated IFR estimates and obtain projections of COVID death counts.
2. Obtain a set of imputed total death counts for each age in the current year.
3. Using this set of imputed values, estimate the first and second levels of parameters in the mortality model by maximising PQL.
4. Determine the remaining parameters on the basis of expert opinions.
5. Simulate N_k paths of k_t , the occurrence times of COVID-alike pandemics and their impacts over a specific forecast horizon on the basis of the parameters estimated and determined in the previous steps.
6. Calculate simulated paths of future mortality.
7. Repeat steps 2–5 for N_p set of imputed values.
8. Repeat steps 1–6 for N_f set of IFR estimates.

8. Conclusion

In this paper, we contribute a stochastic mortality model that uses three levels of parameters to incorporate information from historical mortality data, external forecasts, and expert opinions, respectively. While historical mortality data inform the model about long-term mortality dynamics, external forecasts and expert opinions, which continuously evolve with the pandemic, train the model with the newest knowledge acquired about the pandemic. From the model, insurers and pension plan sponsors can obtain mortality scenarios which would aid them in assessing their mortality/longevity risk exposures.

We find that the commonly used two-stage method for estimating stochastic mortality models cannot disentangle the impact of COVID and fluctuation in long-term mortality improvements. We address this problem with a one-stage estimation method, which simultaneously decomposes mortality into age and period effects and estimates the process for the period effect dynamics.

Historical mortality data and three different sets of external forecasts are collected to estimate the first two levels of parameters. We use multiple imputation to generate age-specific non-COVID death counts in 2020, which are required for parameter estimation but are not yet available as of this writing. Parameter estimates are subject to uncertainty, due to potential errors in the multiple imputation, uncertainty surrounding the external forecasts, and the differences among different available external forecasts. We illustrate how parameter uncertainty can be incorporated into the simulation of mortality paths and demonstrated that the latter two sources contribute the most to the overall parameter uncertainty. If the practitioners ignore the parameter uncertainty in the simulation of future mortality scenarios, they may severely underestimate the risk of their portfolio.

The estimates of the first-level parameters, which intend to capture long-term mortality dynamics, change slightly when different imputed age-specific total death counts in 2020 are used as input; however, given their purpose, the first-level parameters should be insensitive to the imputed 2020 death counts. In future research, one may investigate how this problem may be ameliorated by considering other model structures such as those developed from the CBD model.

The third-level parameters in the model are designated to incorporate expert opinions on the long-term impact of COVID and the possibility of COVID-alike pandemics in the future. In our illustration, we demonstrate how the third-level parameters may turn out to be when different expert opinions are adopted. In the current version of the model, only one set of expert opinions can be incorporated at a time. A possible direction of future research is to elicit opinions from multiple experts and incorporate these opinions into the mortality model simultaneously using a Bayesian approach.

Our study examines the pandemic impact on all-cause mortality. While an analysis of mortality by cause of death may be more informative, it requires more granular data which are difficult to obtain at the early days of a pandemic. Most early epidemiological studies on COVID mortality forecast the number of COVID deaths only, but not the number of deaths by other causes, due to limited data availability. As the pandemic evolves and mortality data by cause of death become available, it is warranted to analyse and model the pandemic impact on mortality by cause of death.

References

- Biggs, A.T. & Littlejohn, L.F. (2020). Revisiting the initial covid-19 pandemic projections. *The Lancet Microbe*, **2**, 91–92.
- Breslow, N.E. & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88** (421), 9–25.
- Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31** (3), 373–393.
- Cairns, A., Blake, D., Kessler, A. & Kessler, M. (2020). The impact of covid-19 on future higher-age mortality. Available online at the address https://www.researchgate.net/publication/341509319_The_Impact_of_Covid-19_on_Future_Higher-Age_Mortality.
- Cairns, A.J., Blake, D., Dowd, K., Coughlan, G.D. & Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41** (1), 29–59.
- Cairns, A.J.G., Blake, D. & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *The Journal of Risk and Insurance*, **73** (4), 687–718.
- Centres for Disease Control and Prevention. (2020). Covid-19 pandemic planning scenarios. Available online at the address <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-archive/planning-scenarios-2020-09-10.pdf> [accessed 22-March-2021].
- Chen, H. & Cox, S.H. (2009). Modeling mortality with jumps: applications to mortality securitization. *The Journal of Risk and Insurance*, **76** (3), 727–751.
- Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C.A.B., Quilty, B.J., Kucharski, A.J., Gibbs, H., Clifford, S., Gimma, A., van Zandvoort, K., Munday, J.D., Diamond, C., Edmunds, W.J., Houben, R.M.G.J., Hellewell, J., Russell, T.W., Abbott, S., Funk, S., Bosse, N.I., Sun, Y.F., Flasche, S., Rosello, A., Jarvis, C.I., Eggo, R.M. & C.C. Working Group. (2020). Age-dependent effects in the transmission and control of covid-19 epidemics. *Nature Medicine*, **26** (8), 1205–1211.
- Delwarde, A., Denuit, M. & Eilers, P. (2007). Smoothing the lee–carter and poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical Modelling*, **7** (1), 29–48.
- Deng, Y., Brockett, P.L. & MacMinn, R.D. (2012). Longevity/mortality risk modeling and securities pricing. *Journal of Risk and Insurance*, **79** (3), 697–721.
- Edwards, M. (2020). Sources of future mortality variation resulting from the pandemic. *Institute and Faculty of Actuaries Longevity Bulletin*, **13**, 35–38.
- Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G. & Dighe, A. (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. Available online at the address <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>.
- Haberman, S. & Renshaw, A. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, **50** (3), 309–333.
- Kuhnert, P., Martin, T. & Griffiths, S. (2010). A guide to eliciting and using expert knowledge in bayesian ecological models. *Ecology Letters*, **13** (7), 900–914.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87** (419), 659–671.
- Li, J.S.-H., Zhou, R., Liu, Y., Graziani, G., Hall, R.D., Haid, J., Peterson, A. & Pinzur, L. (2020). Drivers of mortality dynamics: identifying age/period/cohort components of historical US mortality improvements. *North American Actuarial Journal*, **24** (2), 228–250.
- Lin, Y. & Cox, S.H. (2008). Securitization of catastrophe mortality risks. *Insurance: Mathematics and Economics*, **42** (2), 628–637.
- Liu, Y. & Li, J.S.-H. (2015). The age pattern of transitory mortality jumps and its impact on the pricing of catastrophic mortality bonds. *Insurance: Mathematics and Economics*, **64**, 135–150.
- McCulloch, C.E. & Neuhaus, J.M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, **26** (3), 388–402. doi: [10.1214/11-STS361](https://doi.org/10.1214/11-STS361). Available online at the address <https://doi.org/10.1214/11-STS361>.

- Murray, J.V., Goldizen, A.W., O'Leary, R.A., McAlpine, C.A., Possingham, H.P. & Choy, S.L. (2009). How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? a case study using brush-tailed rock-wallabies petrogale penicillata. *Journal of Applied Ecology*, **46** (4), 842–851.
- Noymer, A. (2011). The 1918 influenza pandemic hastened the decline of tuberculosis in the United States: an age, period, cohort analysis. *Vaccine*, **29** (Supplement 2), B38–B41.
- O'Driscoll, M., Ribeiro Dos Santos, G., Wang, L., Cummings, D.A.T., Azman, A.S., Paireau, J., Fontanet, A., Cauchemez, S. & Salje, H. (2021). Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, **590** (7844), 140–145.
- Oliver, N. (2020). Long covid: the longer-term effects of covid-19. *Institute and Faculty of Actuaries Longevity Bulletin*, **13**, 12–13.
- R.J. Verrall, P.E. (2005). Incorporating expert opinion into a stochastic model for the chain-ladder technique. *Insurance: Mathematics and Economics*, **37** (2), 355–370.
- Ross, A.G., Crowe, S.M. & Tyndall, M.W. (2015). Planning for the next global pandemic. *International Journal of Infectious Diseases*, **38**, 89–94.
- The Human Mortality Database. (2021). Available online at the address <https://www.mortality.org/> [accessed 22-March-2021].
- Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*, 4th edition. Springer, New York. Available online at the address <https://www.stats.ox.ac.uk/pub/MASS4/>. ISBN 0-387-95457-0.
- Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P.G.T., Fu, H., Dighe, A., Griffin, J.T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani, G., Riley, S., van Elsland, S., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly, C.A., Ghani, A.C. & Ferguson, N.M. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, **20** (6), 669–677. doi: [10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- Zhou, R., Li, J.S.-H. & Tan, K.S. (2013). Pricing standardized mortality securitizations: a two-population model with transitory jump effects. *Journal of Risk and Insurance*, **80** (3), 733–774.