Comparison of accuracy between LC model and 4-PFM when COVID-19 impacts mortality structure

Janghoon Choi[1,a]

[a]Korea Insurance Research Institute, Korea

Abstract

This paper studies if the accuracies of mortality models (LC model vs. 4-parametric model) are aggravated if a mortality structure changes due to the impact of COVID-19. LC model (LCM) uses dimension reduction for fitting to the log mortality matrix so that the performance of the dimension reduction method may not be good when the matrix structure changes. On the other hand, 4-parametric factor model (4-PFM) is designed to use factors for fitting to log mortality data by age groups so that it would be less affected by the change of the mortality structure. In fact, the forecast accuracies of LCM are better than those of 4-PFM when life-tables are used whereas those of 4-PFM are better when the mortality structure changes. Thus this result shows that 4-PFM is more reliable in performance to the structural changes of the mortality. To support the accuracy changes of LCM the functional aspect is explained by computing eigenvalues produced by singular vector decomposition

Keywords: Lee-Carter model, 4-parametric factor model, accuracy, mortality structure, reliable

1. Introduction

COVID-19 epidemic is spreading at a rapid rate, getting people afraid of being infected. According to the Korea Disease Control and Prevention Agency (2020.12.30., KDCA), the confirmed and death cases continue to occur and those total cumulative numbers of the countries whose cumulative confirmed cases are over 590,000 are 67,628,461 and 1,536,244 respectively, leading to the total fatality rate of 2.27% as of 2020.12.30. The fatality rates are different by ages, and they increase rapidly for the elderly (Promislow, 2020). In Korea the fatality rates for the elderly over age 80 are approximately 16.4%. The mortality of the country as a whole might change if the confirmed cases are continuously increasing. Moreover, the mortality might not be returned for a short period of time because COVID-19 not only directly causes to death but also indirectly affects the body to death later. For example, people can be damaged by internal injuries such as kidney function declining from COVID-19 virus or aggravate their health status due to not receiving healthcare service in right time by movement restrictions. Therefore, mortality forecasting needs to be performed again by applying COVID-19 effect if it significantly spreads out. To perform forecasting, the accuracies of current mortality models should be evaluated in advance whether they perform well in the structural changes of a mortality.

The most commonly used model for forecasting mortality is Lee-Carter model (LCM) because the model is simple to use and the performance is quite good. There are various research papers regarding LCM and related models such as Lee and Carter (1992), Li and Lee (2005), Lee *et al.* (2016), Park *et al.* (2013), Kang *et al.* (2006), Renshaw and Haberman (2006), Currie (2013), Wiśniowski *et al.*

---

[1] Korea Insurance Research Institute, 38 Gukjegeumyung-ro 6-gil, Youngdeungpo-gu, Seoul 07328, Korea.
  E-mail: james021@gmail.com

(2015), Nigri *et al.* (2019), Li and Li (2017), and much more. However, LCM may not be reliable in forecasting performance to the structural changes of a mortality, but recently developed 4-parametric factor model (4-PFM) might be more reliable in this situation because 4-PFM can flexibly respond to the structural changes by using 4 factors for fitting to the mortality data by age groups. The papers related to the reliability in performance are not many but a few. Li *et al.* (2012) pointed out that mortality forecasts by using LCM would be biased and inaccurate if a mortality index k is not linear. Hunt and Villegas (2015) showed that LCM with cohort effect added is not robust to the change of data or algorithm. Ukert *et al.* (2017) showed that the model Chapman *et al.* (2016) designed didn't reflect the mortality change due to the gun control act of Australia in yr1996. This paper introduces 4-PFM and compares the accuracies with those of LCM, first, by using life-tables, and second, by using 3 scenarios assuming COVID-19 impact and change the mortality structure. I set out the research as follows. In Section 2, two mortality models are explained, focusing more on 4-PFM. Then the accuracy tests are performed for the life-tables in Section 3, and for 3 scenarios in Section 4. Finally I evaluate the models and conclude.

2. Mortality models

2.1. Lee-Carter model

LCM is developed by Lee and Carter in yr1992. In this model log mortality rate is equal to average log mortality rate plus the product of an age dependent sensitivity value and a period dependent trend value plus an error term as in equation (2.1).

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \tag{2.1}$$

where $m_{x,t}$ is the mortality rate of age x in time t, a_x is the average log mortality rate at age x, k_t is a trend factor value of the log mortality rate at time t, b_x represents a sensitivity of age x for k_t, and $\epsilon_{x,t}$ is an error term. b_x and k_t are obtained by using singular value decomposition (SVD) for $\ln(m_{x,t}) - a_x$. Since SVD may lead to numerous solutions, b_x and k_t are standardized such that the sums to be 1 and 0 respectively as in equation (2.2) to obtain a unique solution.

$$\sum_{x=0}^{X} b_x = 1, \sum_{t=t_0}^{T} k_t = 0, \tag{2.2}$$

where X is the age of death, T is the most recent year for the data used. Final forecast result is produced by first fitting a_x, b_x, and k_t to the log mortality data, and then forecasting k_t by analyzing time-series of the past k_t's, then applying the forecasted value of k_t to equation (2.1).

2.2. 4-parametric factor model

4-PFM (Haldrup *et al.*, 2019) is the model by applying Dynamic Nelson-Siegel model which forecasts term structure of interest rates used in financial market to mortality. 4-PFM is expressed in (2.3).

$$\ln(m_{x,t}) = \beta_{0t} + \beta_{1t} e^{-\lambda_1 x} + \beta_{2t} e^{-\lambda_2 (\ln x - \ln k)^2} + \beta_{3t} \left(\frac{x}{N}\right)^{\lambda_3} + \epsilon_{x,t}. \tag{2.3}$$

It is designed such that 4 factors ($\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{3t}$) and corresponding factor loadings are fitted to the log mortality data. Factor loadings are determined by parameters $\lambda_1, \lambda_2, \lambda_3$, and k which are all positive. N is the age of death. β_{0t} is a common factor applying to all ages (Siler 1979), β_{1t} is a factor
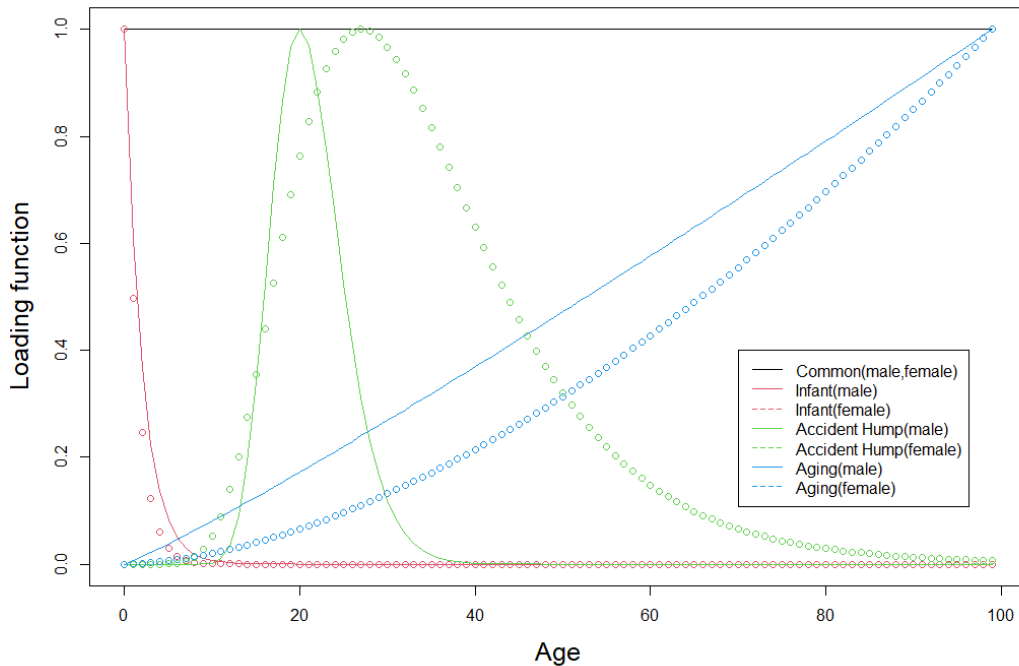
Figure 1: *Shapes of the loading functions by using Korea life-table: yr1983~yr2018.*

reflecting mortality of infants (Rogers and Little 1994), β_{2t} is a factor reflecting mortality from the accident humps in young ages (Heligman and Pollard 1980), and β_{3t} is a factor reflecting mortality rate increase as age increases (Gompertz 1825).

β_{0t} is applied to all ages in common so that the corresponding loading function is 1. The loading function corresponding to β_{1t} decreases very rapidly from age 0 as age increases. The bigger λ_1 is, the more rapidly decreasing the function is. The loading function corresponding to β_{2t} has a bell shape on the central axis of age k. The bigger λ_2 is, the narrower the bell shape of the function is. The loading function corresponding to β_{3t} increases linearly when λ_3 equals 1, is concave when λ_3 is less than 1, and convex when λ_3 is greater than 1. $\epsilon_{x,t}$ is an error term and is assumed to follow iid $N(0, \sigma^2)$.

The average shapes of the loading functions by using the Korea life-tables from yr1983 to yr2018 are in Figure 1. The infant loading function (red) shows that the infant mortality rate. The rate for female infants decreases more rapidly than the one for male infants. The accident hump loading function (green) has a narrower and more symmetric bell shape for male than for female. The accident mortality peak occurs at age around 20 for male and occurs later at age around 30 for female. The ageing loading function (blue) increases linearly for male but, for female, increase slower in young ages and faster in later ages.

The estimation process of 4-PFM in this paper consists of two levels. Level 1 fits the model to the data by least squares method, and level 2 forecasts mortality rates by analyzing time trends of the fitted factors. The estimation methods in level 1 are a little different between for 'fitted accuracy' and 'forecast accuracy'. For level 1 in 'fitted accuracy' test, the parameters are estimated by minimizing the sum of squares for all years at once. It is expressed in (2.4). Then the factors are estimated for

each year given the fixed parameters estimated in (2.4). It is expressed in (2.5).

$$(\lambda_1, \lambda_2, \lambda_3, k) = \arg\min \sum_{t=1983}^{2018} \sum_{x=0}^{T} (y_{xt} - \hat{y}_{xt})^2,$$ (2.4)

$$(\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{3t}) = \arg\min \sum_{x=0}^{T} (y_{xt} - \hat{y}_{xt})^2 \quad \text{given} \ (\lambda_1, \lambda_2, \lambda_3, k),$$ (2.5)

where y_{xt} is the log mortality from the life-table, $\hat{y}_{xt}$ is the estimated value from 4-PFM, x is age, T is age of death, and t is year. However, for level 2 in 'forecast accuracy' test, the parameters are estimated for the most recent year and the factors are estimated for each year given the estimated parameters so that the fitted factors have more reliable time trends. For example, if we have life-tables from yr1983 to yr2018 and perform 'forecast accuracy' test for yr2018, the parameters are estimated for yr2017 life-table as in (2.6).

$$(\lambda_1, \lambda_2, \lambda_3, k) = \arg\min \sum_{x=0}^{T} (y_{x,2017} - \hat{y}_{x,2017})^2.$$ (2.6)

Then, the factors are estimated each year from yr1983 to yr2017 life-tables given the fixed parameters estimated in (2.6). It is expressed in (2.7),

$$(\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{3t}) = \arg\min \sum_{x=0}^{T} (y_{xt} - \hat{y}_{xt})^2 \quad \text{given} \ (\lambda_1, \lambda_2, \lambda_3, k).$$ (2.7)

In level 2, the fitted factor values are forecasted by vector auto-regression (VAR) analysis and then put in the model producing the forecasted mortality rates. The parameters estimated for 'forecast accuracy' test could be used for 'fitted accuracy' test, but the forecast accuracies are not better than by using the parameters estimated by the most recent year (e.g., yr2017).

In Section 3, the results of fitted and forecasted values of the models are shown and the accuracies between 4-PFM and LCM are compared.

3. The accuracy comparison between two models by using Korea life-table

The accuracy implies how models reflect the real data. The accuracies are measured by root mean square error (RMSE) as in equation (3.1),

$$\text{RMSE} = \sqrt{\frac{\sum_{x=0}^{T} (Y_x^m - Y_x^r)^2}{T + 1}},$$ (3.1)

where Y_x^m is the estimated value of the log mortality rate at age x, Y_x^r is the real value of log mortality rate at age x, and T is the age of death.

'Accuracy' is divided into two parts, 'fitted accuracy' and 'forecast accuracy'. The former is the accuracy between real data and the fitted data produced by the models, and the latter is the accuracy between real data and the forecasted data produced by analyzing time trends of the fitted data. Thus 'forecast accuracy' somewhat depends on 'fitted accuracy'.
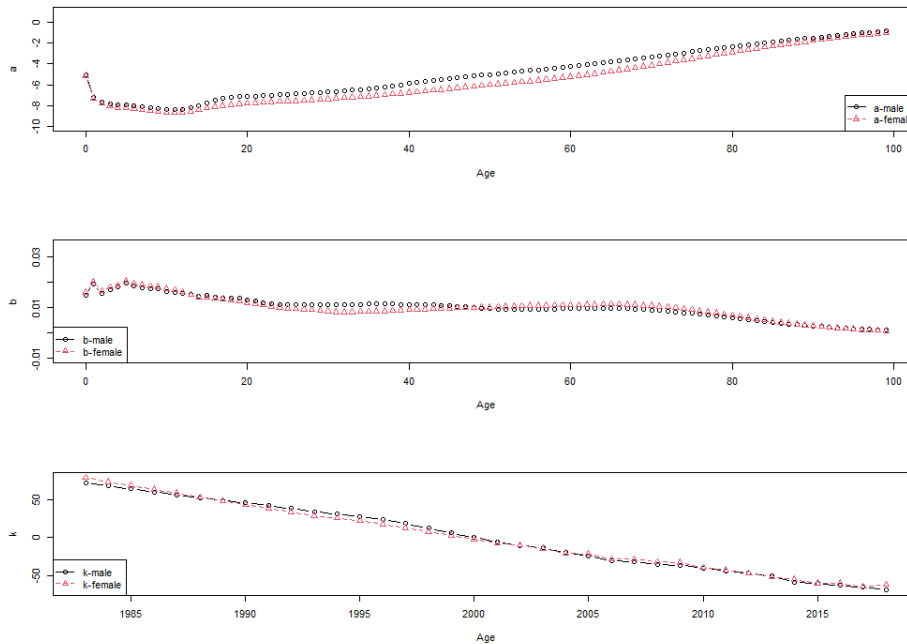
Figure 2: *Fitted variables of LCM.*

3.1. Fitted accuracy

The Korea life-tables from yr1983 to yr2018 are used to test 'fitted accuracy'(Statistics Korea 2020). The life-tables before yr1983 are available but they are not used because they are less reliable. The life-tables have mortality rates from ages 0~100 and the mortality rate at age 100 is fixed to 1, so the mortality rate at age 100 is dropped and the ages 0~99 are used. For 'fitted accuracy'test, the models are fitted. LCM is fitted by using SVD, and 4-PFM is fitted by using least square method.

The fitted variables of LCM are shown in Figure 2. a_x increases with ages for both sexes. a_x of female is located lower than that of male but moves closer to that of male as age increases. b_x moves down slowly with ages and a little faster since age 70 for both sexes. b_x of male is located higher than that of female in ages about 20~50, which means the mortality rates of male for ages 20~50 are more sensitive to k_t than those of female. k_t's for both sexes move down as time passes, and the two lines are nearly overlapped.

For 'fitted accuracy' test 4-PFM is fitted in level 1 as mentioned in Section 2. The limited memory BFGS procedure (L-BFGS-B) via the R package 'optim' is used for the fitting as in Haldrup and Rosenskjold (2019). The average fitted parameters are in Table 1 and the henceforth loading values are shown in Figure 1. λ_1 of female is larger than that of male. It means the mortality rates in female infants change more rapidly than male infants. λ_2 is larger for male, which means the mortality rates by accident are more rapidly changing for male than for female. k is 7 point larger for female than for male, meaning that the death of female by accident occurs in average 7 years later than that of male. λ_3 is larger for female, so the mortality rates for female increase faster than for male in later ages.

Next, the time trends of the 4 fitted factors are shown in Figure 3. First, the common factor β_{0t} decreases linearly both for male and female, so overall mortality rates decrease as time goes

Table 1: Fitted parameter values of 4-PFM

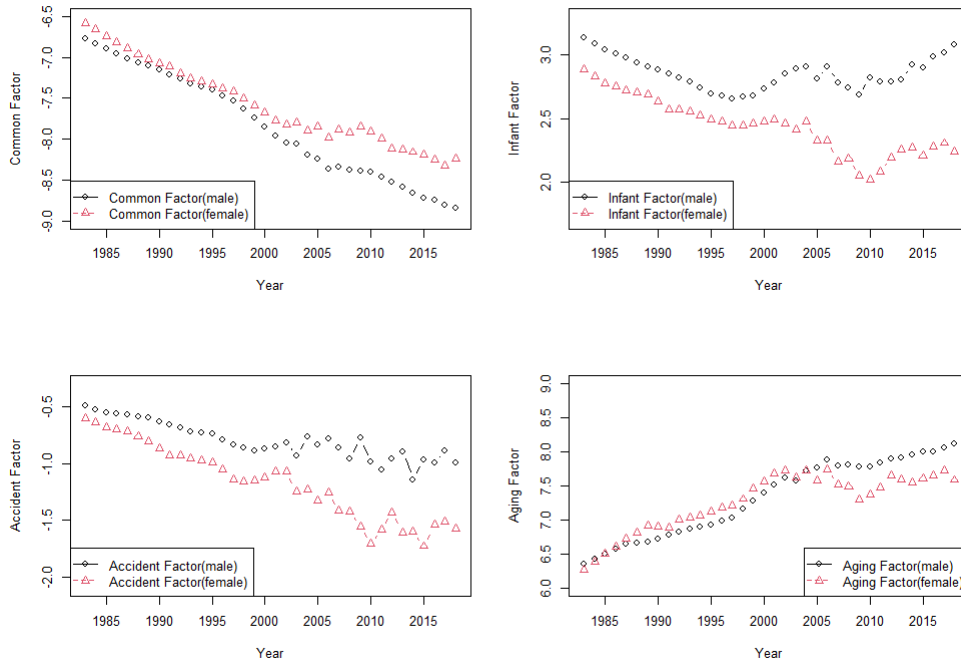| Parameter	λ_1	λ_2	λ_3	k
Male	0.5	13	1.1	20
Female	0.7	3	1.7	27

Figure 3: *Trends of 4 fitted factors of 4-PFM.*

Table 2: 'fitted accuracy' (RMSE): 4-PFM vs. LCM, averaged from yr1983 to yr2018

Male				Female		
4-PFM(1)	LCM(2)	(1) / (2)		4-PFM(1)	LCM(2)	(1) / (2)
0.104	0.055	1.891		0.130	0.079	1.646

'Fitted accuracy' is measured by RMSE.

by. However the decreasing speed of male since yr2000 is faster than female. The infant factor β_{1t} smoothly decreases for both sexes until yr2000 and then turns to be volatile since then. The gap between male and female is larger since yr2000. The accident factor β_{2t} smoothly decreases for both sexes until yr1997 and becomes volatile since that year, The gap between the two lines is wider as time passes, which means male is more exposed to the death by accident than female as time passes. Finally, the aging factor β_{3t} increases for both sexes but increasing speed decreases since yr2000. Thus it is implied that the aging factor's effect on the mortality has decreased since yr2010. The volatility of β_{3t} for female increases since that year.

The result of the average 'fitted accuracy' from yr1983 to yr2018 is shown in Table 2. The average RMSE of 4-PFM from yr1983 to yr2018 is larger than that of LCM for both sexes. Thus 'fitted accuracy' is better for LCM. The average RMSE of 4-PFM is 1.9 times larger than LCM for male,

Table 3: Result of Johansen test of 4-PFM for test 1 (yr2018)

Number of cointegration	Test		Significance level		
	Male	Female	10%	5%	1%
$r \leq 3$	1.83	1.83	6.50	8.18	11.65
$r \leq 2$	7.59	6.07	15.66	17.95	23.52
$r \leq 1$	16.84	21.61	28.71	31.52	37.22
$r = 0$	40.20	62.91	45.23	48.28	55.43

R is number of cointegration. Life-tables of yr1983~yr2017 are used.

and 1.6 times larger for female.

3.2. Forecast accuracy

To test 'forecast accuracy', forecasted results are produced by time trend analysis for the fitted data estimated in level 1 and differences between real log mortality rates and forecasted results are computed. Since future mortality rates are not known, a point of time A in the past is assumed as a current point, and forecast is performed based on the assumed current point A.

'Forecast accuracy' can be affected by the data characteristics at some specific point of time. For example, the accuracy would not be good though the forecast by the time trend analysis is accurate if the data at some specific point to measure the accuracy may not follow the past time trend, or the accuracy might be good by luck if the data at some specific point are fortunately close to the forecasted data though the time trend analysis is not accurate. To reduce coincidence, two tests are performed based on two different point of time. Test 1 assumes yr2017 as a current point of time and forecasts mortality rates at yr2018 and computes RMSE between forecasted data and real data at yr2018. Test 2 assumes yr2013 as a current point of time and forecasts mortality for the next 5 years and computes RMSE between the 5 year-forecasted data and the 5 year-real data from yr2014 to yr2018.

By using 'auto.arima' function in R package for test 1, the appropriate ARIMA model for k_t in LCM is estimated to be ARIMA(0, 1, 0) for male, and ARIMA(1, 2, 1) for female. For 4-PFM, VAR analysis regarding 4 factors are performed to check autocorrelation and cointegration. Portmanteau test for autocorrelation analysis and Johansen test for cointegration analysis are used. Refer to Pfaff Bernhard (2008) for detailed explanation. The result of the autocorrelation analysis shows that the vector of 4 factors for male has 3rd order autocorrelation, and that for female has 1st order autocorrelation. The cointegration analysis result is in Table 3. There is no cointegration effect for male because the test value when the number of cointegration (r) is 0 is not rejected on the significance level of 10%. However there is one cointegration relation for female because the test value when $r = 0$ is rejected on the significance level of 1%.

The result of 'forecast accuracy' for yr2018 by the time trend analysis is shown in Table 4. The result shows that the performance is better for LCM for both sexes. The ratio of RMSE of 4-PFM over LCM is 1.5 for male, and 1.1 for female. It is abnormal that 'forecast accuracy' of LCM for male is 0.082 whereas 'fitted accuracy' is 0.084 which is larger than 'forecast accuracy' because 'forecast accuracy' includes 'fitted accuracy'. Thus RMSE of 'forecast accuracy' might be computed to be small by luck.

Now we move to test 2 which is to test the forecast for yr2014~yr2018. The time trend analysis for LCM shows that ARIMA(2, 1, 0) is appropriate for male, and ARIMA(1, 2, 1) is appropriate for female. For 4-PFM, there exist 1st order autocorrelation for both sexes, and the result of cointegration analysis shows that there is no cointegration effect for male but there is one cointegration relation for female as in Table 5. The time trend analysis results of test 2 are different from those of test 1. It may

Table 4: Accuracy (RMSE) for test 1 (yr2018): 4-PFM vs. LCM

Accuracy	Male			Female		
	4-PFM(1)	LCM(2)	(1) / (2)	4-PFM(1)	LCM(2)	(1) / (2)
Fitted accuracy(a)	0.095	0.084	1.130	0.130	0.093	1.394
Forecast accuracy(b)	0.124	0.082	1.518	0.181	0.158	1.143

The values of 'fitted accuracy' are those in yr2017 in Table 2.

Table 5: Result of Johansen test of 4-PFM for test 2 (yr2014~yr2018)

Number of cointegration	Test		Significance level		
	Male	Female	10%	5%	1%
$r \leq 3$	0.03	0.07	6.50	8.18	11.65
$r \leq 2$	6.01	4.15	15.66	17.95	23.52
$r \leq 1$	18.63	16.07	28.71	31.52	37.22
$r = 0$	35.77	75.31	45.23	48.28	55.43

R is number of cointegration. Life-tables of yr1983~yr2013 are used.

Table 6: Accuracy (RMSE) for test 2 (yr2014~yr2018): 4-PFM vs. LCM

Accuracy	Male			Female		
	4-PFM(1)	LCM(2)	(1) / (2)	4-PFM(1)	LCM(2)	(1) / (2)
Fitted accuracy(a)	0.083	0.049	1.695	0.133	0.073	1.824
Forecast accuracy(b)	0.120	0.102	1.168	0.200	0.155	1.289

*The values of 'fitted accuracy' are those in yr2013 in Table 2.
**'Forecast accuracy' is the average RMSE from yr2014 to yr2018.

imply that the data period used is not long enough.

The results of 'forecast accuracy' for yr2014~2018 by the time trend analysis are shown in Table 6. The ratio of RMSE of 4-PRM over LCM is 1.2 for male, and 1.3 for female. Thus test 2 result shows that the performance of LCM is also better than that of 4-PFM as in test 1.

4. The accuracy comparison between two models by COVID-19 scenarios

4.1. Fatality rates caused by COVID-19 and scenario set-up

As of December 30, 2020, the cumulative number of confirmed case of COVID-19 in Korea was about 115.3 per 100,000 people, which was significantly lower than 5,765.1 in US, 992.0 in Singapore, 3,477.2 in UK, and 2,020.3 in Germany. However, it is hard to forecast how long COVID-19 exists, and how much more it spreads out in the future (Kim, 2020).

COVID-19 does not affect the fatality rates equally for all ages. The fatality rates increase rapidly with age. According to KDCA, the cumulative number of confirmed cases and death from COVID-19 as of December 30, 2020 were 59,773 and 879, respectively, for all ages, 2,958 and 486 for ages over 80, 4,702 and 250 for ages of 70~79, and 9,458 and 103 for ages of 60~69.

Although the difference between the fatality rates by COVID-19 and the mortality rates of the life-tables has been calculated in a short term, it can last for a long time or even increases because COVID-19 might not be controlled in a short period of time and furthermore the virus can damage body function such as lungs, which may not cause deaths in a short period but many years later.

In this section, the accuracies between the two models are compared by scenarios assuming that the mortality rates are impacted by COVID-19. The accuracies for the scenarios are also compared with those for the life-tables to test which model is more reliable to the impact.

To set up the scenarios, the ratios of fatality rates of the confirmed people over the mortality rates

Table 7: Ratios of the two death rates by age: confirmed people of COVID-19 over life-table

Age	Mortality rate of life-table (1)	Fatality rate of confirmed people (2)	Ratio (2/1)
60~69	0.7131%	1.09%	1.53
70~79	2.1403%	5.32%	2.49
80+	8.4077%	16.43%	1.95

Source: Life-table (yr2018)-Statistics, Korea (2020) KDCA (Korea Disease Control and Prevention Agency 2020.12.30).

of the yr2018 life-table are produced and applied to the life-tables. The reason why the ratios are applied to the life-tables and the fatality rates are not used directly is because the fatality rates are only the current rates but we need time trend rates. By applying the ratios to the life-tables in different years, we can obtain the scenarios of time trend data. The ratios computed by using the two death rates by age are in Table 7. The fatality rates of the confirmed people provided by KDCA exist only from the age above 30 so it is assumed that there is no fatality rate below the age 30. Also the fatality rates of the ages below 60 are lower than those of the life-table so only those of the ages over 60 are applied to the life-tables. The overall fatality rates of Korea for the ages over 60 are 1.5~2.5 times higher than those of Korea life-table. The ratio of the two death rates of the age 70's is 2.49 which is the highest, and then 80's and 60's in this order.

Three scenarios are set up by applying Table 7. For all of the three scenarios, the life-tables of yr1983~yr2018 are used, and the fatality rates caused by COVID-19 are set up to start in yr2001. Scenario 1 has the same impact on all the life-tables from yr2001 to yr2018, but scenario 2 and scenario 3 have different impacts on each life-table as follows.

- Scenario 1: The ratios of 1.53 for age 65, 2.49 for age 75, and 1.95 for age 90 are multiplied to the life-tables from yr2001 to yr2018 identically, then the impacted mortality rates are smoothed.

- Scenario 2: The ratios are set up with decrease as time passes, that is, the same ratios as in scenario 1 are multiplied to the yr2001 life-table, then the ratios reduce gradually untill 1 is multiplied to the yr2018 life-table. Thus the scenario mortality in yr2018 is the same as that of the yr2018 life-table.

- Scenario 3: The ratios are multiplied to the life-tables such that they increase from yr2001 to yr2009 in which the ratios are peaked to the same as in scenario 1, then from yr2009 turn to decrease until yr2018 where the scenario mortality rates are the same as in the yr2018 life-table.

Figure 4 shows the log mortality rates by scenario. The plots above are the log mortality rates of male and the below are those of female. The mortality curves of three years which are yr2001, yr2009, and yr2018 are displayed to show the time trends. The black curves are those of the life-tables and the red, green, and blue color curves are those of the scenarios for yr2001, yr2009, and yr2018, respectively. All the 3 scenario curves are folded with the life-table curves below the ages of 60 because the scenario mortality rates are not impacted before the ages of 60.

In scenario 1, the log mortality rates decrease as the year passes as in the life-tables because the same ratios are multiplied from yr2001 to yr2018. Thus there is no time trend distortion. However, there are age trend distortions as you can see the shapes of the impacted curves. When compared with the life-tables, the impacted curves are located above the life-tables since the ages of 60.

In scenario 2, the scenario curves since the ages of 60 are located above the life-tables in yr2001 and yr2009, but return to the life-table in yr2018. Since the multiplied ratios reduce, there are some time trend distortions such as a large gap between yr2009 and yr2018 scenario curves. There are also the age trend distortions as in scenario 1.

In scenario 3, the levels of the scenario curves are more complicated than scenario 1 and 2. As in scenario 2, the scenario curves of yr2001 and yr2009 are located above the life-tables in yr2001
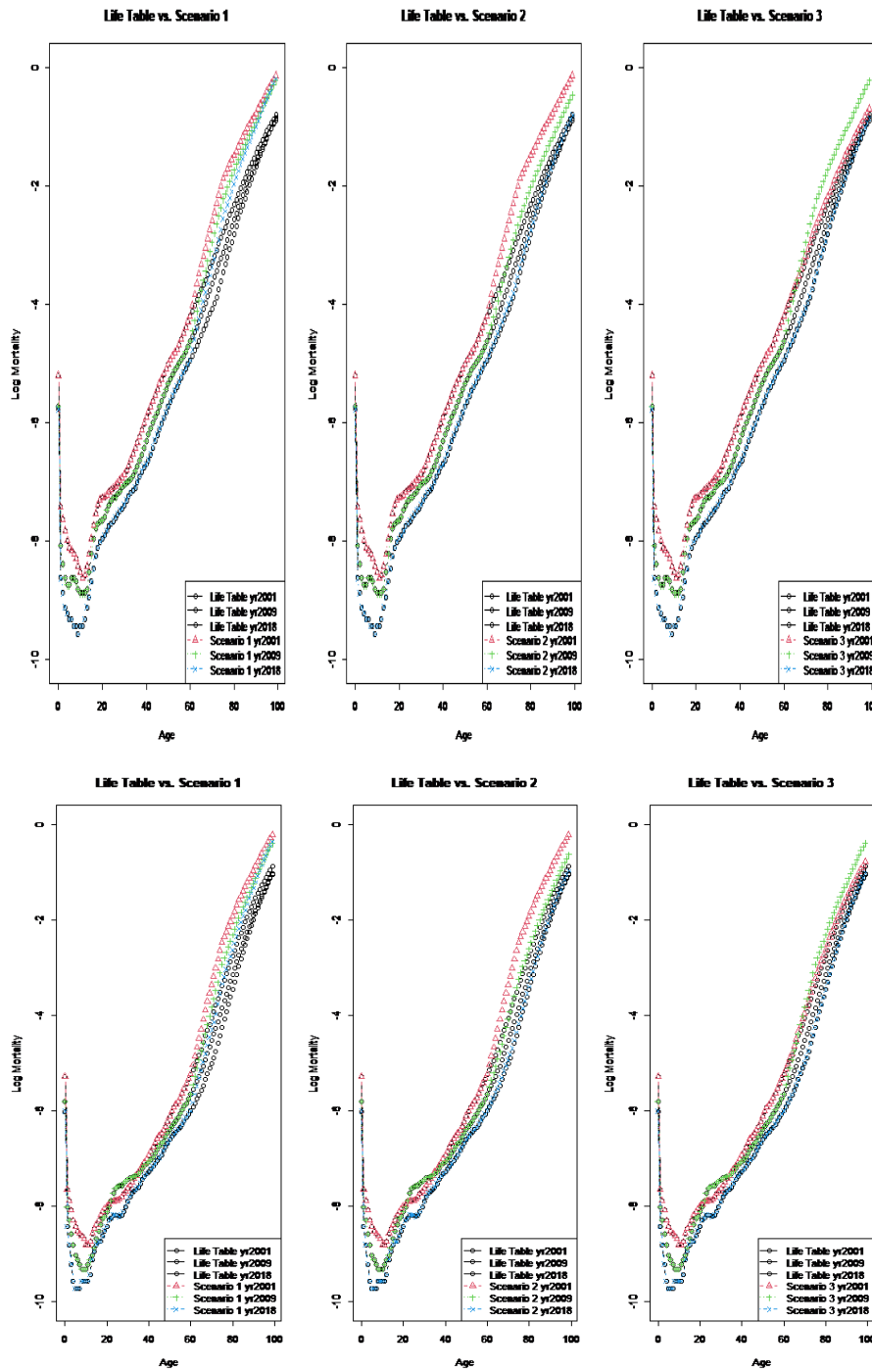
Figure 4: *Log mortality by scenario (above: male, below: female).*

Table 8: 'fitted accuracy' by scenario (RMSE): 4-PFM vs. LCM

Scenario	Male			Female		
	4-PFM(1)	LCM(2)	(1) / (2)	4-PFM(1)	LCM(2)	(1) / (2)
Life-table	0.105	0.058	1.808	0.131	0.083	1.570
Scenario 1	0.163	0.121	1.347	0.195	0.133	1.466
Scenario 2	0.138	0.153	0.902	0.171	0.169	1.012
Scenario 3	0.135	0.130	1.038	0.167	0.137	1.219

The results are average RMSE from yr1983 to yr2018.

Table 9: RMSE ratio over life-table for 'fitted accuracy'

Scenario	Male		Female	
	4-PFM(1)	LCM(2)	4-PFM(1)	LCM(2)
Scenario 1	1.552	2.086	1.489	1.602
Scenario 2	1.314	2.638	1.305	2.036
Scenario 3	1.286	2.241	1.275	1.651

The numbers are calculated by RMSE of the scenarios divided by RMSE of the life-table.

and yr2009 but the scenario curve of yr2018 is folded on the yr2018 life-table. The levels between the two scenario curves of yr2001 and yr2009 are different depending on ages. The yr2001 scenario curve is located above the yr2009 scenario curve as in scenario 1 and 2 when the age is below 67, but the level is reversed when the age is above 67. It is because the multiplied ratios are different between the two years. They are larger for yr2009 than for yr2001. Thus there are distinctive time trend distortions such as level reversion between yr2001 and yr2009 in scenario 3. There are also the age trend distortions as in both scenario 1 and 2.

4.2. Accuracy by scenario

The estimation process of the models for the scenarios is the same as in section 3 and the variable estimation results are in Appendix.

In Section 4.2, the results of the accuracies between two models are compared by scenario. First, 'fitted accuracy' and 'forecast accuracy' are shown, and second, the reason why the accuracies of the scenarios are changed is explained.

First, 'fitted accuracy' is shown in Table 8. The accuracies are aggravated for both models in all the 3 scenarios than in the life-table. For 4-PFM, the accuracies are worst in scenario 1, while, for LCM, they are worst in scenario 2. When compared between the two models, the accuracies in the scenarios are still better for LCM except scenario 2 for male.

However, in terms of the level of aggravation in 'fitted accuracy' as the mortality structure changes, 4-PFM is less aggravated than LCM as you can see in Table 9. The numbers in Table 9 represent the RMSE ratios of the scenarios over the life-table. The RMSE ratios for 4-PFM are 1.3~1.6 for male, which are less than those for LCM. For female, the ratios for 4-PFM are 1.3~1.5, which are also less than those for LCM.

Next, 'forecast accuracy' is compared in two separate tests. Test 1 is performed by forecasting to yr2018, and test 2 is by forecasting to yr2014~yr2018 as in Section 3. The results are shown in Table 10. When the two models are compared, the both tests show that 4-PFM is better than LCM in scenario 2 and 3 as the ratios are less than 1 except in scenario 1. Compared among the scenarios, the accuracies with scenario 1 and scenario 3 from the both tests are worse than those with scenario 2 for 4-PFM, while the accuracies with scenario 2 and scenario 3 are worse than scenario 1 for LCM. When compared between sexes, the overall accuracy is better for male than for female, for the both

Table 10: 'forecast accuracy' by scenario (RMSE): 4-PFM vs. LCM

Scenario		Male			Female		
		4-PFM(1)	LCM(2)	(1) / (2)	4-PFM(1)	LCM(2)	(1) / (2)
Life-table[a]	yr2018[b]	0.124	0.082	1.518	0.181	0.158	1.143
	5 yr avg[c]	0.120	0.102	1.168	0.200	0.155	1.289
Scenario 1	yr2018	0.186	0.152	1.224	0.246	0.193	1.275
	5 yr avg	0.201	0.189	1.063	0.239	0.220	1.086
Scenario 2	yr2018	0.124	0.294	0.422	0.186	0.347	0.536
	5 yr avg	0.171	0.349	0.490	0.198	0.386	0.513
Scenario 3	yr2018	0.135	0.356	0.379	0.187	0.398	0.470
	5 yr avg	0.230	0.344	0.669	0.256	0.376	0.681

[a]The results from Life-tables are from Section 3. [b]Computed from forecasted value of yr2018. [c]Computed from forecasted value of yr2014~yr2018.

Table 11: RMSE ratio over Life-table for 'forecast accuracy'

Scenario		Male		Female	
		4-PFM(1)	LCM(2)	4-PFM(1)	LCM(2)
Scenario 1	yr2018[a]	1.500	1.854	1.359	1.222
	5 yr avg[b]	1.675	1.853	1.195	1.419
Scenario 2	yr2018	1.000	3.585	1.028	2.196
	5 yr avg	1.425	3.422	0.990	2.490
Scenario 3	yr2018	1.089	4.341	1.033	2.519
	5 yr avg	1.917	3.373	1.280	2.426

[a]RMSE ratio of scenario1 over the life-table for forecasting yr2018.
[b]Average RMSE ratios of scenario1 over the life-table for forecasting yr2014~yr2018.

models.

In terms of the level of aggravation in 'forecast accuracy' as the mortality structure changes, 4-PFM is less aggravated than LCM as you can see in Table 11. For male, the RMSE ratios of the scenarios over the life-table are approximately 1.0~1.9 for 4-PFM, which are less than those for LCM. For female, the ratios are 1.0~1.4 for 4-PFM, which are also less than those for LCM except scenario 1.

Thus we can conclude that 4-PFM is more reliable in performance to the structural changes than LCM based on the both 'fitted accuracy' and 'forecast accuracy' tests.

Second, the reason why the performance of LCM is aggravated when mortality structure is distorted is explained as follows. LCM applies SVD to the model estimation. The SVD method decomposes the log mortality matrix, after subtracted by average log mortality vector $(a_0, a_1, \ldots, a_X)$, into three matrices which are age dependent matrix, scaling diagonal matrix, and period dependent matrix. LCM uses only the largest singular value on the scaling diagonal matrix and its corresponding vectors to reduce dimensions. $(b_0, b_1, \ldots, b_X)$ is the vector in the age dependent matrix, corresponding to the largest singular value, and $(k_{t_0}, k_{t_1}, \ldots, k_T)$ is the product of the largest singular value and its corresponding vector in the time dependent matrix. Thus the performance of LCM depends on the accuracy of dimension reductions. To have accurate dimension reductions, the columns or the rows of the matrix should have similar patterns, and those of the log mortality matrix do in a normal situation, that is, the mortality rates always increase as the age increases and the increasing pattern is similar for different times, and also the mortality rates decrease as time passes and the decreasing speed is gradual and not volatile. Thus the log mortality matrix can be decomposed into age related, and time related matrices with each having similar patterns of vectors so that the two vector's $((b_0, b_1, \ldots, b_X)$ and $(k_{t_0}, k_{t_1}, \ldots, k_T))$ can replace their corresponding matrices quite accurately by SVD. However, if the mortality structure is distorted such that the patterns of the vectors in the matrices are not similar,

Table 12: Accuracy of decomposition of LCM: *Prop*1

Scenario	Male	Female
Life-table	0.985	0.970
Scenario 1	0.935	0.919
Scenario 2	0.911	0.877
Scenario 3	0.925	0.915

the dimension reduction technique does not work well. We can guess how accurate the dimension reduction is by computing '*Prop*1' in equation (4.1),

$$\text{Prop1} = \frac{d_1^2}{\sum_{i=1}^{r} d_i^2}, \tag{4.1}$$

where r is equal to or smaller than the smaller value between the age of death and the number of periods, d_i's are singular values so that d_i^2's are eigenvalues, and d_1 is the largest singular value. *Prop*1 is less than 1. If *Prop*1 is close to 1, the accuracy of the decomposition is high. The values of *Prop*1 by scenario are shown in Table 12. As we expect, *Prop*1's for using the life-tables are closer to 1 than for using the scenarios. When we compare among the 3 scenarios, the decompositions work well in scenario 1, 3, 2 in this order for male, and scenario 3, 1, 2 in this order for female.

Moreover, since the inaccurately fitted values of k_t are used to forecast by time trend analysis in the scenarios, the 'forecast accuracy' of LCM is aggravated further.

On the other hands, 4-PFM uses 4 factors with each factor fitted for the mortality rates by some range of ages. Since the factors would be adjusted for the distorted parts, the distortion of mortality structure does not affect significantly the model's performance.

5. Conclusion

In this paper the accuracy between LCM and 4-PFM is compared, assuming that COVID-19 impacts the mortality structure. The accuracy tests were performed in two levels, one with 'fitted accuracy' and the other with 'forecast accuracy'.

Both accuracy tests showed that the performance of LCM was better than 4-PFM when the life-table data were used, but was not always better when the mortality structure was changed. The performance of LCM was better for 'fitted accuracy' test, but 4-PFM worked better for forecast accuracy' test. The level of the aggravation also showed that 4-PFM was relatively more reliable than LCM.

The performance of LCM was explained by defining *Prop*1 which was the ratio of the largest eigenvalue over the sum of all eigenvalues. From the results of the various analyses, we concluded that LCM worked well in a normal situation, but was vulnerable to the change of the mortality structure, when compared with 4-PFM.

Currently the most popular mortality model is LCM. However, the performance of LCM is not always reliable. We cannot say that COVID-19 has influenced the mortality rates such that their structure is changed so far, because the number of confirmed people and the number death from COVID-19 hold small proportions over the population. It is possible, however, that if COVID-19 spreads out or various mutants occur, the mortality structure could be changed. Therefore, more in-depth analyses regarding the performance of mortality models are required.

Finally, 4-PFM needs to be improved. The model is more reliable than LCM, but the accuracy in a normal situation is not good enough compared with LCM. Designing new PFM adding more factor loadings to current 4-PFM to increase the accuracies could be the next research project.

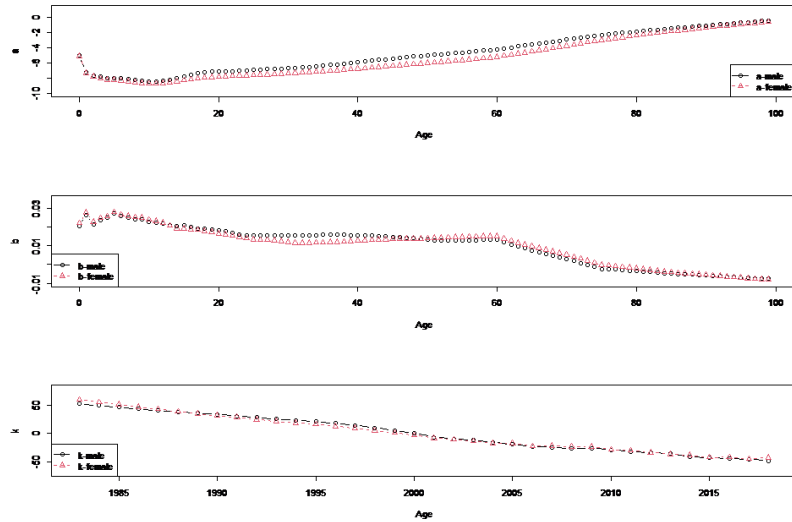## **Appendix A: Variable estimation of the models by scenario**

Scenario 1

Figure A.1: *Estimation of LCM: a_x, b_x, k_t.*

Table A.1: Parameter estimation of 4-PFM

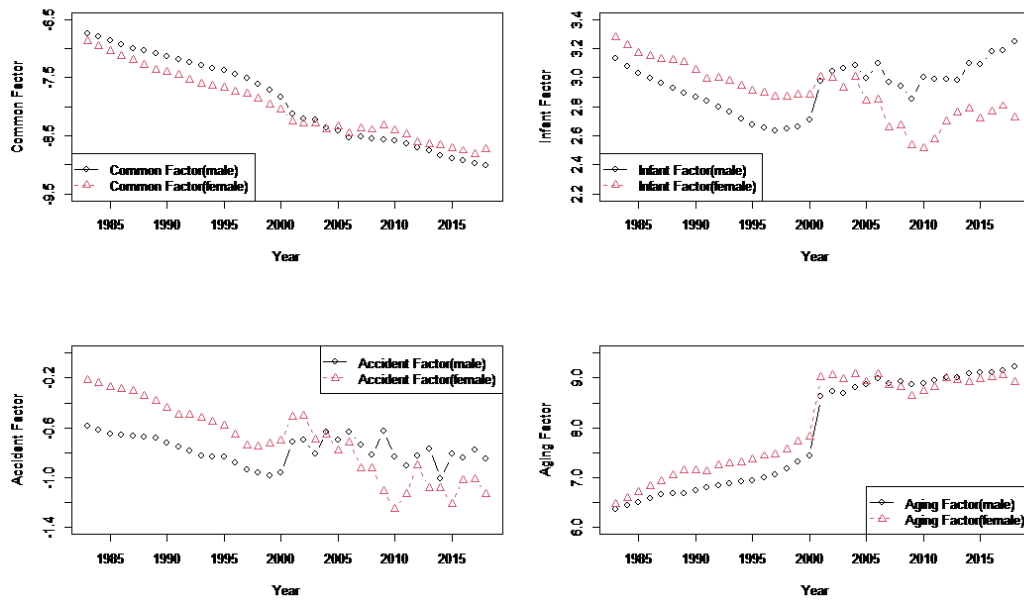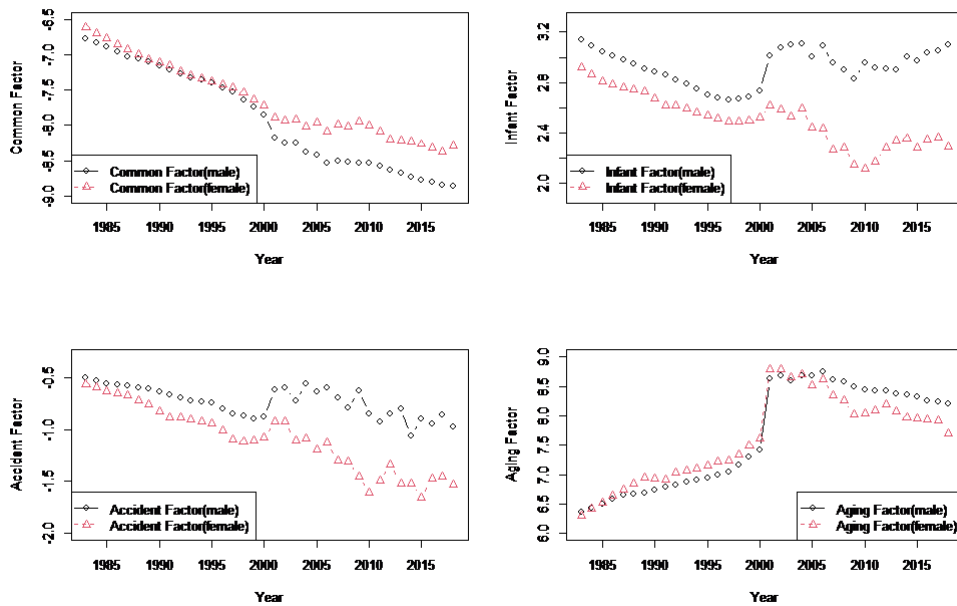| Sex	λ_1	λ_2	λ_3	k
Male	1.1	13.9	1.3	10.8
Female	1.0	15.3	1.8	10.6

Figure A.2: *Time trends of 4-PFM factors*

Scenario 2
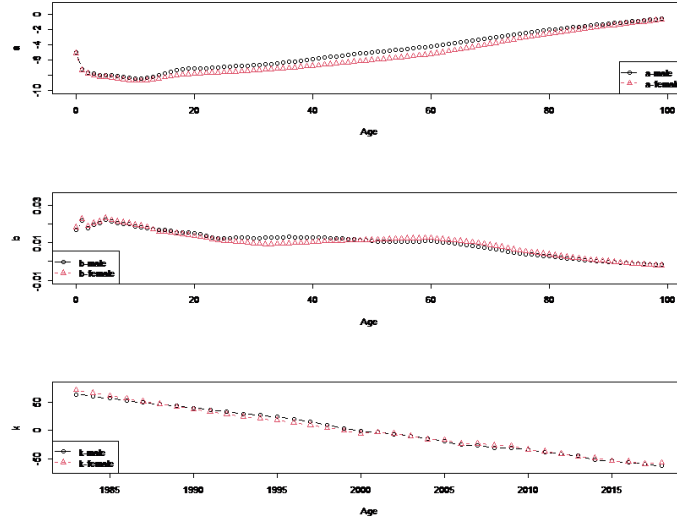
Figure A.3: *Estimation of LCM: a_x, b_x, k_t.*

Table A.2: Parameter estimation of 4-PFM

Sex	λ_1	λ_2	λ_3	k
Male	1.1	13.3	1.3	10.7
Female	1.7	14.7	1.8	10.5

Figure A.4: *Time trends of 4-PFM factors.*

Scenario 3
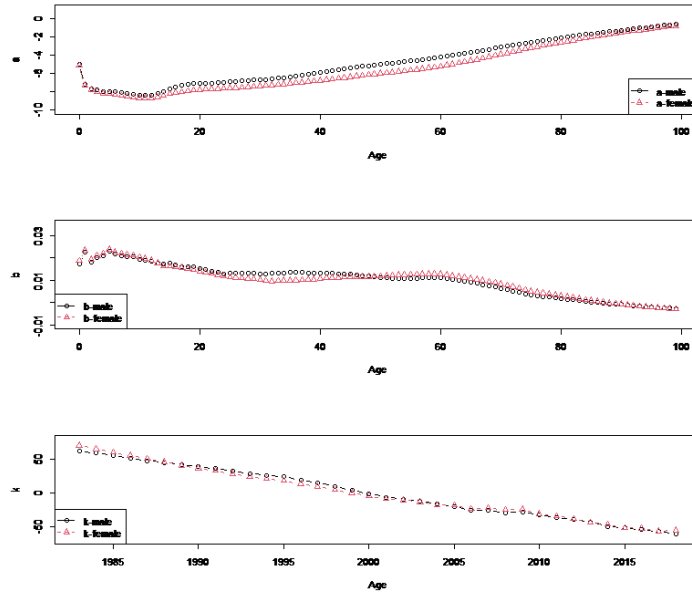
Figure A.5: *Estimation of LCM: a_x, b_x, k_t*

Table A.3: Parameter estimation of 4-PFM

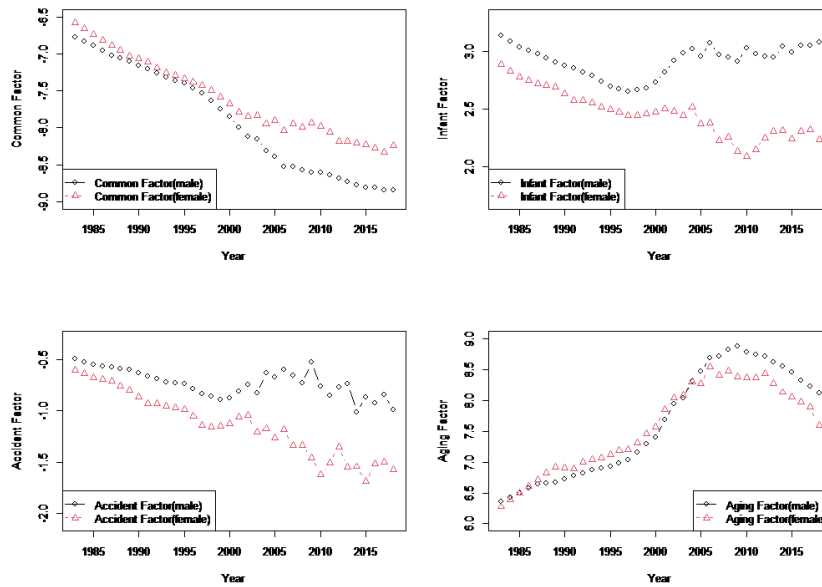| Sex	λ_1	λ_2	λ_3	k
Male	1.1	13.1	1.3	10.7
Female	1.6	14.1	1.8	10.5

Figure A.6: *Time trends of 4-PFM factors.*

References

Chapman S, Alpers P, and Jones M (2016). Association between gun law and intentional firearm deaths in Australia, 1979–2013. *Journal of the American Medical Association*, **316**, 291–299.

Currie L (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model, *Statistical Modelling*, **13**, 69–93.

Gompertz B (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Society of London*, **115**, 513–583.

Haldrup N and Rosenskjold PT (2019). A parametric factor model of the term structure of mortality, *Econometrics*, Retrieved from July 9th,2019,doi:10.3390/econometrics7010009.

Heligman L and Pollard JH (1980). The age pattern of mortality, *Journal of the Institute of Actuaries*, **107**, 49–80.

Hunt A and Villegas AM (2015). Robustness and convergence in the Lee-Carter model with cohort effects, *Insurance: Mathematics and Economics*, **64**, 186–202.

Kang JC, Lee DS, and Sung JH (2006). A study on the methods for forecasting mortality considering longevity risk, *The Journal of Risk Management*, **17**, 153–178.

Kim NS (2020). Current situation and task of COVID-19, *ISSUE & FOCUS*, Korea Institute for Health and Social Affairs, ISSN 2092–7117.

Lee HS, Baek CR, and Kim JH (2016). A modified Lee-Carter model based on the projection of the skewness of the mortality, *Korean Journal of Applied Statistics*, **29**, 41–59.

Lee RD and Carter LR (1992). Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, **87**, 659–671.

Li H and Li J (2017). Optimizing the Lee-Carter approach in the presence of structural changes in time and age patterns of mortality improvements, *Demography*, **54**, 1073–1095.

Li J, Chan WS, and Cheung SH (2012). Structural changes in the Lee-Carter mortality indexes: detection and implications, *North American Actuarial Journal*, **15**, 13–31.

Li N and Lee R (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method, *Demography*, **42**, 575–594.

Nigri A, Levantesi S, Marino M, Scognamiglio S, and Perla F (2019). A deep learning integrated Lee-Carter model, *Risks*, **7**.

Park YS, Jang SW, and Kim SY (2013). VECM-LC model for forecasting mortality in Korea, *Survey Research*, **14**, 19–47.

Pfaff B (2008). VAR, SVAR and SVEC models: implementation within R package vars, *Journal of Statistical Software*, **27**, 1–32.

Promislow EL(2020). A geroscience perspective on COVID-19 mortality, *The Journals of Gerontology*.

Renshaw AE and Haberman S (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors, *Insurance: Mathematics and Economics*, **38**, 556–570.

Rogers A and Little JS (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedule, *Mathematical Population Studies*, **4**, 175–195.

Siler W (1979). A competing-risk model for animal mortality, *Ecology*, **60**, 750–757.

Statistics Korea (2020). Life-table, Retrieved from August 30, 2020, https://kosis.kr/statisticsList /statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ZTITLE&parmTabId=M_01_01#SelectSt atsBoxDiv.

Ukert B, Andreyeva E, and Branas CC (2017). Time series robustness checks to test the effects of the

1996 Australian firearm law on cause-specific mortality, *Journal of Experimental Criminology*.

Wiśniowski A, Smith WF, Bijak J, Raymer J, and Forster JJ (2015), Bayesian population forecasting: extending the Lee Carter method, *Demography*, **52**, 1035–1059.