Université Catholique de Louvain

2021-2022

LACTU-2030
Life insurance project: static and dynamic mortality modeling

Students

Maxime Druez (8754-21-00)

Hugo Léal (1847-21-00)

Professor

Donatien Hainaut

December 6, 2021

Contents

1 Introduction

This work aims to implement, discuss and compare several statistical methods seen in class for adjusting static or dynamic life tables to real mortality data. In section 2, we will focus first on static mortality modeling. This framework relies on the death observations of any peculiar year that are interpreted as one possible realization of a random variable that does not evolve over time. Of course, this approach neglects any dynamic effect that could modify the mortality over time, such as a reduction of hazard rates, an increase in life expectancy or any cohort effect. Dynamic or prospective mortality tables wish to go beyond those artefacts and have become a standard in mortality modeling, namely through the R-package `StMoMo` (for stochastic mortality modeling). As such, the prospective approach is further explored in section 3 of this report.

2 Static mortality modeling

2.1 Introduction to static mortality modeling

If D_x denotes the observed number of people passing away at age x (during the year of interest for static modeling) in a population counting L_x individuals at age x, we saw in class how the raw $\hat{q}_x = \frac{D_x}{L_x}$ was an unbiased estimator for the one-year death probability q_x at age x. It was also the log-likelihood estimator of q_x and its variance was shown to be inversely proportional to the size L_x of the population at age x, namely $V(\hat{q}_x) = \frac{q_x p_x}{L_x}$.

Anyway, $\hat{q}_x$ is only an estimator of the true q_x and it can exhibit large fluctuations, especially at large age x (low L_x) where the quality of the estimator is reduced (higher variance). The latter fluctuations in $\hat{q}_x$ cannot be related to a real trend in the mortality and one rather wishes to isolate a smoother model closer to the true curve of decease probabilities q_x.

In order to do so, two main paradigms were seen in class: smoothing techniques or parametric models. The former are studied in section 2.2 through a specific example: the Whittaker Henderson (WH) procedure. The latter are envisioned in sections 2.4 to 2.8 where one wishes to fit a peculiar logit model to raw $\hat{q}_x$. In this parametric framework and for this particular logit model, we will compare the adjustments entailed by a mean square error approach or by a log-likelihood optimisation (under the Gaussian assumption).

Finally, we will illustrate in sections 2.3 and 2.7 how adjusted mortality tables can help in the computation of practical quantities. The results obtained from the WH smoothing will be used to price simple life insurance products whereas the outputs from the parametric approach will enable us to compute life expectancies at various ages.

2.2 Whittaker Henderson smoothing

For years 2008 and 2018, for male and female Belgian populations (of age x $>$ 18y.), life tables were recovered from the moodle website (source mortality.org). Those tables contain in particular the raw $\hat{q}_x$ and the L_x of interest for the smoothing procedure (resp. in columns labelled "qx" and "lx" of the database). In those four cases (male/female in 2008/2018), we wish to smooth the raw $\hat{q}_x$ by the WH smoothing technique.

We can recall that the WH smoothing appears as a trade-off between the quality of the fit and the smoothness of the modeled probability curve. The smoothing parameter h gives the importance granted to the smoothness compared to the quality of the fit. If $q_x^s(h)$ denotes the smooth one-year death probability at age x for parameter h, the WH method chooses $q_x^s(h)$ in order to minimize the quantity $F(q_x^s(h)) + h.S(q_x^s(h))$. The F-quantity,

$$F(q_x^s(h)) = \sum_x w_x (q_x^s(h) - \hat{q}_x)^2$$

measures the quality of the fit. The weights $w_x = \frac{L_x}{L_0}$ are taken proportional to L_x to grant more importance to ages where the quality of the estimator $\hat{q}_x$ is higher (lower variance due to larger population). The S-quantity,

$$S(q_x^s(h)) = \sum_x ((\Delta^2 q_x^s(h))_x)^2$$

measures the smoothness of the curve through the discrete difference operator (that was limited here to the second order).

As explained in class, the WH procedure can be rephrased in a matrix formalism. The smooth curve $q^s(h)$ is then given by

$$q^s(h) = (W + hK_2'K_2)^{-1}W\hat{q}$$

where W is the diagonal matrix of the weights and K_2 is the matrix associated to the second order difference operator (with 1 on the main diagonal and on the second upper diagonal, with -2 on the first upper diagonal and with zeros elsewhere).

In practice, we only performed the smoothing and the quality of the fit discussion for ages x $>$ 18y. for which the Cochran criterion was satisfied:

$$L_x\hat{q}_x \geq 5 \qquad\qquad L_x(1-\hat{q}_x) \geq 5$$

It essentially amounts to exclude large ages x with not enough individuals (too low L_x), as noted in tab. 1. For male populations, with higher mortality, the rejection of the Cochran criterion is more prompt to be seen at lower ages.

	Female 2008	Female 2018	Male 2008	Male 2018
Age min	19	19	19	19
Age max	108	108	106	107
Number of ages m	90	90	88	89

Table 1: Ages x $>$ 18y. for which the Cochran criterion is satisfied and number of ages m satisfying those conditions

For those ages satisfying the Cochran criterion, the matrix WH formalism was implemented in the R-code provided in appendix C. The smoothing parameters h retained in the four cases are given in tab. 2.

	Female 2008	Female 2018	Male 2008	Male 2018
h	3.35	4.28	28	34.9
S	116.97	116.98	114.67	115.76
$\chi^2_{1-2.5\%,m-1}$	116.99	116.99	114.69	115.84

Table 2: Smoothing parameters h (elicited with a sensitivity analysis and with the test of goodness-of-fit) and results for the test of goodness-of-fit (confidence $\alpha = 2.5\%$). Indeed, the test statistics S is lower than $\chi^2_{1-\alpha,m-1}$ such that the fit is not rejected.

Elicitation of h

Provided the Cochran criterion is satisfied, $\hat{q}_x$ has nearly a normal distribution with mean q_x and variance $\frac{q_x p_x}{L_x}$. We can then envision a test for goodness-of-fit. For ages x where the Cochran criterion is satisfied, under the H_0 assumption $q_x = q_x^s(h)$ (i.e. the fit is good), the quantity

$$S = \sum_x L_x \frac{(q_x^s(h) - \hat{q}_x)^2}{q_x^s(h)(1-q_x^s(h))}$$

should behave asymptotically as a χ^2 with (m-1) degrees of freedom (where m represents the number of ages on which the smoothing is performed). For a confidence level α, the H_0 assumption must be rejected if the test statistics S is above $\chi^2_{1-\alpha,m-1}$ (percentile $1-\alpha$ of the χ^2 with (m-1) degrees of freedom).

If we increase h, we grant more importance to the smoothness of the modeled curve w.r.t. the quality of the fit. As such, when h is increased, the test statistics S is worsened in the sense it becomes larger. Eventually, when h is increased too much, the test statistics S could exceed $\chi^2_{1-\alpha,m-1}$ and the H_0 assumption should be rejected. In other words, our fit would become unacceptably poor for a too high h value.

Within this idea of trade-off between smoothness and quality of the fit, the h values given in tab. 2 were elicited in order to maximize the test statistics S, under the constraint that S must remain below $\chi^2_{1-\alpha,m-1}$ for a given confidence level α. In practice, we used $\alpha = 2.5\%$. In other words, we took the maximal h (the maximal smoothness) that one can allow not to reject the proposed fit, at a confidence level $\alpha = 2.5\%$.

Results and analysis
Fig. 1 gives the results of the WH procedure for the male population in 2018. The raw $\hat{q}_x$ and the smooth curve $q_x^s(h)$ obtained by WH smoothing are plotted in semi-log scales. Indeed, the WH curve is smoother than the raw estimator although it remains a qualitative fit of the raw curve. Similar results were obtained for male population in 2008 and for female populations in 2008 and 2018 but those are provided in appendix A.

Let us note that the curves have the expected behaviour. The decease probability is quite low around x = 20-30y. and then it increases almost linearly in log scale up to high ages. On our smooth curves, we do not observe a bump due to accidents around ages x = 20-25y., except maybe for female populations (especially in 2008 cf. appendix A). This could illustrate the efficiency of prevention campaigns.

At very high ages ($\geq 90y.$), we observe a downward inflection of the trend for decease probability, both in raw data and in smooth curves. However, as noted in fig. 1, for male in 2018, the smooth curve (in red) starts to underestimate a bit the raw observations (in blue) at very high ages. This edge effect possibly comes from a lower statistics at high ages, due to a lower population L_x, especially for men having a higher mortality. Indeed, this effect is hardly seen in female plots given in appendix A. However, even for men in 2018 (cf. fig. 1), this remains quite a marginal observation since we limited ourselves to ages where the Cochran criterion was satisfied (which essentially means that L_x was already high enough). Though, if it had to become a problem, one could further reduce the number of ages on which to perform the smoothing and prefer an extrapolation of the model for very high ages. This is not done in the present document for succinctness of the text.

Finally, in fig. 1, the smooth curve for female population in 2018 was also added in order to illustrate the higher mortality for men as already mentioned several times.

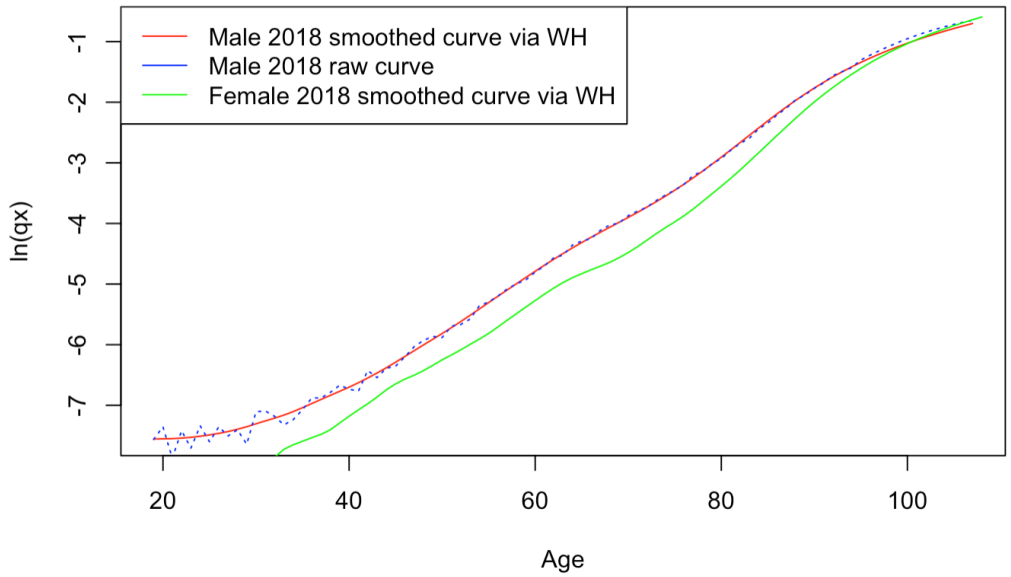

Figure 1: One-year death probability: raw $\hat{q}_x$ and smoothed curve $q_x^s(h)$ obtained by WH smoothing

2.3 Pricing of simple life insurance products

Interpreting now the smooth $q_x^s(h)$ as the true q_x, the results from the previous section can be used in order to price several simple life insurance products.

Term life insurance
For a woman or a man, of age x= 30y., in 2008 or in 2018, we can compute the net single premium of a term life insurance (duration n= 20y. and death capital C= 1Mio €) under the assumption that $i = 0\%$. Since the actualisation factor v reduces to the unity for $i = 0\%$, it is simply given by:

$$\Pi = C._{20}A_{30} = C.\sum_{k=0}^{19} v^{k+0.5}{}_kp_{30}.q_{30+k} = C.\sum_{k=0}^{19} {}_kp_{30}.q_{30+k}$$

In this formula, for the one-year death probability q_{30+k} we simply use the smooth results for male/female in 2008/2018 depending on the case we are. Of course, ${}_0p_{30} = 1$. The remaining survival probabilities ${}_kp_{30}$ (if $k \geq 1$) can also be computed from smooth q_x as follows :

$$ {}_kp_{30} = p_{30}.p_{31}...p_{30+k-1} = (1 - q_{30}).(1 - q_{31})...(1 - q_{30+k-1}) $$

The resulting net single premiums are gathered in tab. 3. Since the mortality is higher for male populations, it is not surprising to have a bigger premium for a man compared to a woman for a contract guaranteeing death benefits, anything else being equal. Even if we worked under a static framework, the raw data of 2008/2018 we relied on to establish smooth curves were affected by dynamic effects. The data we used should thus translate a reduction of mortality over time. Establishing a static model on data from 2008 or 2018 should thus one way or another be affected by the reduction of mortality over time. It is thus not surprising to have a reduction of premiums over time for contracts guaranteeing death benefits, anything else being equal.

	Female 2008	Female 2018	Male 2008	Male 2018
Π [€]	20,317.88	17,022.14	36,657.79	26,973.7

Table 3: Net single premiums for term life insurance (n=20, C=1Mio €, $i = 0\%$), for a woman or a man of age x= 30y., using smooth results of 2008 or 2018.

Life annuity

For a woman or a man, of age x= 65y., in 2008 or in 2018, we can compute the net single premium of a life annuity (annual payment of C=10 000 €) under the assumption that $i = 0\%$. Since the actualisation factor v reduces to the unity for $i = 0\%$, it is simply given by:

$$ \Pi = C.\ddot{a}_{65} = C.\sum_{k=0}^{\infty} v^k {}_kp_{65} = C.\sum_{k=0}^{\infty} {}_kp_{65} $$

Again, ${}_0p_{65} = 1$ and the remaining survival probabilities ${}_kp_{65}$ (if $k \geq 1$) can also be computed from smooth q_x as follows :

$$ {}_kp_{65} = p_{65}.p_{66}...p_{65+k-1} = (1 - q_{65}).(1 - q_{66})...(1 - q_{65+k-1}) $$

In practice, in the formula of the net single premium, the summation up to the infinity was replaced by a summation up to the maximal age (minus 65y.) until which we had performed the smoothing (i.e. the maximal age (minus 65y.) where the Cochran criterion was satisfied cf. tab. 1). Possibly, one could properly extrapolate our smoothing model up to higher ages and extend the summation to the maximal human lifespan. This is not done in the present document, for succinctness of the text. Besides, the marginal contribution of each supplementary term in the summation becomes less and less important at high ages.

The resulting net single premiums are gathered in tab. 4. Since the mortality is lower for female populations, it is not surprising to have a bigger premium for a woman compared to a man for a contract guaranteeing life benefits, anything else being equal. Even if we worked under a static framework, the raw data of 2008/2018 we relied on to establish smooth curves were affected by dynamic effects. The data we used should thus translate a reduction of mortality over time. Establishing a static model on data from 2008 or 2018 should thus one way or another be affected by the reduction of mortality over time. It is thus not surprising to have an increase of premiums over time for contracts guaranteeing life benefits, anything else being equal.

	Female 2008	Female 2018	Male 2008	Male 2018
Π [€]	211,360.4	221,047.6	176,004.3	189,333.7

Table 4: Net single premiums for life annuity (annual payment of C=10 000 €, $i = 0\%$), for a woman or a man of age x= 65y., using smooth results of 2008 or 2018.

2.4 Parametric modeling: preamble

For this section 2.4 and until section 2.8, we will focus on male data for the year 2018 and for ages ranging from 30y. to 109y. We wish now to fit the following parametric model

$$q_x = f_\theta(x) = \frac{exp(a + b.x^c + d.x.I_{x\geq 70})}{1 + exp(a + b.x^c + d.x.I_{x\geq 70})}$$

to raw $\hat{q}_x$, where $\theta = (a, b, c, d)$ is a vector of four real parameters whose values have to be adjusted. In other words, the model we wish to fit can be rewritten as

$$lg(q_x) = a + b.x^c + d.x.I_{x\geq 70}$$

where lg is the logit function $lg(q_x) = ln(\frac{q_x}{1-q_x})$.

More exactly, since we will work under the gaussian assumption for the log likelihood optimisation in section 2.6, we will restrict ourselves to ages where the Cochran criterion is satisfied. Indeed, under the gaussian assumption, we will assume that $\hat{q}_x$ is nearly normally distributed

$$\hat{q}_x \sim N(q_x = f_\theta(x); \sigma_x^2(\theta) = \frac{f_\theta(x)(1 - f_\theta(x))}{L_x})$$

In order to do so, we must assume to have enough data at disposal. In other words, the Cochran criterion must be satisfied on the range of ages on which we will perform the fit. Considering tab. 1, the maximal age where the Cochran criterion is satisfied for male in 2018 is 107y. As such, for log likelihood optimization under the gaussian assumption, we will perform the fit for ages between 30y. and 107y. (instead of 109y. that is proposed in the statement). Since the results of mean square error and log likelihood optimizations will have to be compared, not to introduce a bias in the data selected, the fit through mean square error optimization will also be performed on the same range 30y. to 107y. (instead of 109y. that is proposed in the statement). The results we will obtain could be extrapolated up to higher ages but this is not considered in the present document for succinctness of the text.

As a preamble, we must first check that the proposed model makes sense on the range of ages 30y. to 107y. that we want to fit. It is natural to plot $lg(\hat{q}_x)$ (logit of raw data) to see whether or not it makes sense to develop a model of the type $lg(q_x) = a + b.x^c + d.x.I_{x\geq 70}$. This was done in fig. 2.

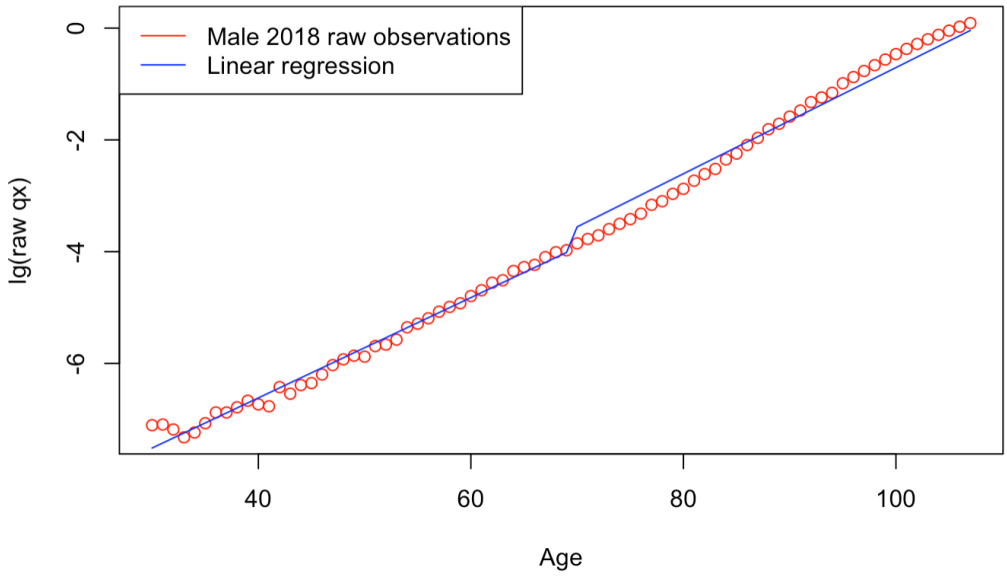

Figure 2: For male data in 2018, logit of raw $\hat{q}_x$ as a function of age and linear regression $a^{lg} + b^{lg}.x^{c^{lg}} + d^{lg}.x.I_{x\geq 70}$

In this figure, we see that the logit of observations are almost linearly aligned from x=30y. to x= 69y. As such, a model of the type $lg(q_x) = a+b.x^c$ with c close to 1 should make sense on this range of ages. Actually, we performed the linear regression of $lg(\hat{q}_x)$ (on a model of the type a+b.x) for this range of ages, entailing the parameters a^{lg} and b^{lg} given in tab. 5 for $c^{lg} = 1$.

In fig. 2, we also see that the logit of observations are almost linearly aligned from x=70y. to x= 107y. As such, a model of the type $lg(q_x) = a + b.x^c + d.x$ with c close to 1 should make sense on this range of ages. Simply, we offer an additional degree of freedom in the model, to better fit the observations at high ages. Through the additional parameter d, we allow a slight modification of the slope of the logit at ages above 70y. Actually, considering a^{lg}, b^{lg} and c^{lg} previously found, we performed the linear regression of $lg(\hat{q}_x) - a^{lg} - b^{lg}.x$ (on a model of the type d.x) for this range of ages , entailing the d^{lg} parameter given in tab. 5.

The linear regressions found by this procedure are also plotted in fig. 2. As such, the proposed model seems to be suitable on the range of ages 30y. to 107y. that we wish to fit. This conclusion would have to be properly assessed through a suitable adequacy test, but this is not done in the present document for succinctness of the text.

a^{lg}	b^{lg}	c^{lg}	d^{lg}
-10.2	0.0897	1	0.00527

Table 5: Initial values for the parameters of the logit model found by linear regression on the logit of raw observations. By virtue of Jensen's inequality, they can only constitute an initial approximation for the parameters since they lead to a biased model.

At this point, we must insist on the fact that parameters a^{lg}, b^{lg}, c^{lg} and d^{lg} found by this procedure cannot be considered as suitable for our model. Indeed, as seen in class, the logit function is increasing and concave (at least between 0 and 0.5). By virtue of Jensen's inequality, we thus have

$$E(lg(\hat{q}_x)) \leq lg(E(\hat{q}_x)) = lg(q_x)$$

This means that the linear regression we performed on observed $lg(\hat{q}_x)$ can only deliver a biased model underestimating the real $lg(q_x)$. This is why the parameters we obtained were indexed by the lg notation, to recall that they can only lead to a biased model.

However, this preamble section and the approach we followed were not without bringing value to the discussion. First of all and as previously mentioned, even if it should be properly formalized through a suitable adequacy test, we demonstrated that the proposed model seems adequate on the range of ages we wish to fit. Secondly, and this might be even more important, the parameters we found by this regression, even if not suitable for the model, can at least offer an initial approximation for the parameters $\theta = (a, b, c, d)$ (whose exact values remain to be found). Indeed, the following sections will attempt to find $\theta = (a, b, c, d)$ through a mean square error or log likelihood optimization procedure. This will involve a numerical optimization problem, solved through an iterative method requesting an initial approximation for the parameters. The mean square error or log likelihood functions we will have to optimize exhibit a rather rough behaviour, with potentially many local optima. To converge to the global optimum and not to be stuck erroneously in a wrong local optimum, it will be important to start the iterations with a not so bad initial approximation of the parameters. As such, the vector $(a^{lg}, b^{lg}, c^{lg}, d^{lg})$ (obtained in this preamble section) will be used as initial value for the subsequent optimization problems of the next sections aiming at finding better values for the parameters $\theta = (a, b, c, d)$. In particular, the degree of freedom on parameter c will be relaxed for the optimizations, whereas we imposed $c^{lg} = 1$ in this preamble section.

2.5 Parametric modeling: mean square error (MSE)

In the mean square error (MSE) paradigm, the vector of parameters $\theta = (a, b, c, d)$ is elicited in order to minimize the distance between the raw $\hat{q}_x$ and the modeled probability $q_x = f_\theta(x)$ in the quadratic sense on the ranges of ages to fit. In other words, θ^{MSE} minimizes the quantity

$$MSE = \sum_x (f_\theta(x) - \hat{q}_x)^2$$

where the summation acts on ages x= 30y. to x= 107y. where we wish to perform the fit.

At this stage, we must mention that the MSE approach does not involve any weights in front of the quadratic errors $(f_\theta(x) - \hat{q}_x)^2$. All ages have the same importance in the MSE quantity. In particular, we do not account for the size of the population L_x. We do not weight the quadratic errors by the inverse of the variance of the estimator $\hat{q}_x$. When we will compare the results of the MSE paradigm with the ones of the log likelihood under the gaussian approximation, we will see the shortfalls of forgetting the weights related to the variance in the reasoning. As we will see, under the gaussian approximation, the log likelihood optimization essentially reduces to a MSE procedure including weights inversely proportional to the variance of $\hat{q}_x$. The objective of the project statement is precisely to study the benefits brought by those weights in the log likelihood framework. This is what will be done in section 2.8 comparing the results of pure MSE (without weights) and of log likelihood under the gaussian approximation (essentially MSE with weights).

Starting with the initial values of parameters $(a^{lg}, b^{lg}, c^{lg}, d^{lg})$ of tab. 5, the R-function `optim` (using the BFGS quasi-Newton method) was used in order to find θ^{MSE} minimizing the MSE quantity. The optimal parameters in the MSE sense are gathered in tab. 6. The resulting one-year death probabilities $q_x = f_\theta(x)$ for θ^{MSE} (on the range of ages x= 30y. to x= 107y.) are plotted in normal and semi-log scales in fig. 3. The discussion about those curves and the comparison with the log likelihood results are to be found in section 2.8.

a^{MSE}	b^{MSE}	c^{MSE}	d^{MSE}
-10.21437	0.05899	1.10456	0.00136

Table 6: Optimal parameters for the logit model in the MSE sense found by `optim`.

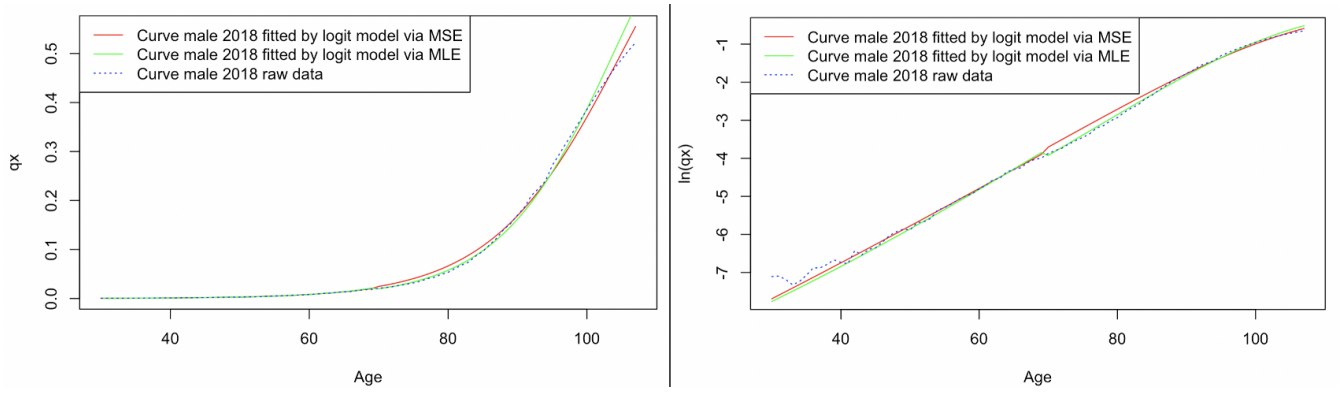

Figure 3: For male data in 2018, in normal and semi-log scales, for the range of ages x= 30y. to x= 107y. (where the Cochran criterion is satisfied), the raw $\hat{q}_x$ as well as the fitted one-year death probabilities $q_x = f_\theta(x)$ (logit model) obtained after the MSE or after the MLE (gaussian approximation) procedures

2.6 Parametric modeling: log-likelihood (MLE) under the gaussian approximation

In what follows, the log likelihood maximization paradigm will be referred to as MLE and we will sometimes forget to mention we work under the gaussian assumption. The latter hypothesis is however essential if one wishes to use the simplified procedure that will be exposed here below. A more general binomial approach was described in class but it might be less robust, it was leading to more difficult optimization problems, it gave less intuitive log-likelihood expressions. It was not presented in this report for succinctness of the text. Anyway, assuming for simpleness that $\hat{q}_x$ is normally distributed (on the range of ages where the Cochran criterion is satisfied), the MLE paradigm elicits the vector of parameters θ^{MLE} maximizing the following log likelihood

$$ln(L(\theta)) = \sum_x \left[-ln(\sigma_x(\theta).\sqrt{2\pi}) - \frac{1}{2}\frac{(f_\theta(x) - \hat{q}_x)^2}{\sigma_x^2(\theta)} \right]$$

The summation acts on ages x= 30y. to x= 107y. where we wish to perform the fit. For numerical considerations, we optimize the log likelihood and not the likelihood itself. Since the R-function `optim` performs a minimization, we will rather minimize the opposite of the log likelihood.

In case the optimization problem is unstable numerically, one could replace $\sigma_x^2(\theta)$ by its estimator $\hat{\sigma}_x^2 = \frac{\hat{q_x}(1-\hat{q_x})}{L_x}$. Yet, this was not even necessary in the present case and performing this manipulation entailed in fact parameters less optimal regarding the log likelihood.

As previously mentioned, under the gaussian approximation, the MLE framework essentially reduces to MSE weighted by the inverse of the variance of the estimator $\hat{q_x}$. Indeed, except for the terms in $ln(\sigma_x(\theta).\sqrt{2\pi})$ (that are even independent of θ if we replace $\sigma_x^2(\theta)$ by its estimator), minimizing $-ln(L(\theta))$ essentially reduces to minimizing the sum of quadratic errors $(f_\theta(x) - \hat{q}_x)^2$ weighted by the inverse of $\sigma_x^2(\theta)$. As such, comparing MSE (of section 2.5) and MLE under the gaussian approximation (of this section 2.6) essentially amounts to discuss the influence of the weights inversely proportional to the variance. This discussion will be conducted in section 2.8.

Starting with the initial values of parameters $(a^{lg}, b^{lg}, c^{lg}, d^{lg})$ of tab. 5, the R-function `optim` (using the BFGS quasi-Newton method) was used in order to find θ^{MLE} minimizing $-ln(L(\theta))$. The optimal parameters in the MLE sense are gathered in tab. 7. The resulting one-year death probabilities $q_x = f_\theta(x)$ for θ^{MLE} (on the range of ages x= 30y. to x= 107y.) are plotted in normal and semi-log scales in fig. 3. The discussion about those curves and the comparison with MSE results are to be found in section 2.8. Confidence intervals could be derived for the parameters estimated *via* MLE, in accordance to the procedure described in class. The goodness-of-fit could also be properly assessed through a suitable test. The latter two objectives were not presented in the current document, not to burden too much the discussion.

a^{MLE}	b^{MLE}	c^{MLE}	d^{MLE}
-9.83391	0.02707	1.27578	-0.00266

Table 7: Optimal parameters for the logit model in the MLE sense (gaussian approximation) found by `optim`.

2.7 Parametric modeling: life expectancy

For a man in 2018, using the fitted $q_x = f_\theta(x)$ obtained in the MSE or MLE formalisms, we can compute the life expectancy at ages x= 40, 50 or 60y. Under the approximation of a piecewise constant mortality rate (constant rate between year t and t+1 for any t natural), we demonstrated in class that the life expectancy at age x can be approached by

$$\mathring{e}_x = \sum_{t=1}^{w-x} t.{}_tp_x.q_{x+t}$$

where w denotes the maximal human lifespan. In practice, w was replaced here by the maximal age for which we had performed the fit of q_x, i.e. x= 107y. (maximal age for which the Cochran criterion was satisfied and for which the gaussian assumption made sense). Anyway, the successive terms have less and less marginal importance for increasing ages. Yet, our model of q_x could be extrapolated to higher ages if precision requirements had to become even more stringent and if the sum had to be extended. This is not done in the present report.

In this formula, for the one-year death probabilities q_{x+t}, we simply used the fitted $q_{x+t} = f_\theta(x+t)$ obtained in the MSE or MLE formalisms. The survival probabilities ${}_tp_x$ (for $t \geq 1$) can also be computed from fitted q_x as follows:

$${}_tp_x = p_x.p_{x+1}...p_{x+t-1} = (1-q_x).(1-q_{x+1})...(1-q_{x+t-1})$$

The resulting life expectancy at x=40, 50 and 60y. for the MSE and MLE cases are gathered in tab. 8. For the sake of comparison, the life expectancy mentioned in life tables (column "ex") of the moodle website (source mortality.org) is also given in tab. 8. Those values are discussed and compared in section 2.8.

	MSE	MLE (under gaussian assumption)	Life tables (moodle)
$\mathring{e}_{40}$	39.01662	39.87862	40.45
$\mathring{e}_{50}$	29.66068	30.48172	31.11
$\mathring{e}_{60}$	20.92011	21.71155	22.38

Table 8: For a man in 2018, at x= 40, 50 or 60y., life expectancy computed from MSE or MLE fitted probabilities or extracted from life tables of the moodle website

2.8 Parametric modeling: comparison of MSE and MLE results

Log likelihood in the MSE and in the MLE cases
In tab. 9, we gathered the log likelihoods of the observations, obtained with the parameters estimated *via* MSE (cf. tab. 6) or *via* MLE (cf. tab. 7). Fortuitously, the log likelihood in the MLE case is higher. Indeed, per definition, the MLE framework tries to find parameters maximizing the log likelihood. Any other method for parameters estimation can only result in less optimal parameters regarding the log likelihood, if the optimizations were correctly performed. This is what is observed here.

	MSE	MLE (under gaussian assumption)
Log Likelihood $ln(L(\theta))$	-471.9599	294.3483

Table 9: Log likelihoods of the observations, obtained with the parameters estimated *via* MSE (cf. tab. 6) or *via* MLE (cf. tab. 7)

It is only when the weights (inversely proportional to the variance of the estimator $\hat{q}_x$) are included in front of the quadratic errors $(f_\theta(x) - \hat{q}_x)^2$ that a log likelihood optimization is performed (under the gaussian assumption). MLE (under the gaussian assumption) looks like MSE but with those weights related to the variance. When we forget the weights, as we did in the pure MSE approach (simply minimizing the sum of quadratic errors), the log likelihood is not optimized in general: the resulting log likelihood can only be sub-optimal for the estimated parameters *via* MSE.

A discussion about curves in fig. 3
We will now discuss the fitted one-year death probabilities $q_x = f_\theta(x)$ (logit model) obtained after the MSE or after the MLE (gaussian approximation) procedures. We recommend the reader to focus on the plot in semi-log scale in fig. 3 to have an illustration of the present discussion.

In the MSE framework, parameters are estimated in order to minimize the sum of quadratic errors $(f_\theta(x) - \hat{q}_x)^2$ (without any weights related to the variance). As such, all ages have an equal importance in the optimization procedure, regardless of the size of the population L_x. If the model performs poorly on ages where $\hat{q}_x$ is high (i.e. at high ages), the related quadratic error terms at those ages are unacceptably high and sub-optimal. If the model performs poorly on ages where $\hat{q}_x$ is low (i.e. at low ages), the related quadratic error terms at those ages would anyway remain quite limited. Indeed, we expect the quadratic errors terms to be anyway of the order of magnitude of $\hat{q}_x^2$ or even less.

As such, in the MSE framework, it is worse to perform poorly at high ages than at low ages. The MSE framework will naturally try to reach a better fit at high ages. This is indeed what is observed in fig. 3 (red curve for MSE). The MSE model tries to reach the best fit possible at high ages. In fact, the MSE model benefits from the parameter d (extra degree of freedom for $x \geq 70y.$) in order to reach an excellent fit around age x= 69y. and around x= 107y. The discontinuity is here positive ($d^{MSE} \geq 0$ cf. tab. 6). We have a positive jump that allows to reach an excellent fit for the highest ages below 70y. (so around x= 69y) and for the highest ages above 70y. (so around x= 107y). Indeed, in fig. 3, the red curve for MSE is really close to the blue curve of raw $\hat{q}_x$ for x= 69y. or for x above 100y. It is in that range of high ages that a poor model would have a tremendous impact on quadratic errors $(f_\theta(x) - \hat{q}_x)^2$ since $\hat{q}_x$ is already high. Of course, this excellent fit at high ages comes with a price. The MSE fit is less qualitative in the range of intermediate or low ages. For the lowest ages below 70y. (so for x=30, 40y.) and for the lowest ages above 70y. (so for x=70, 80y.), the red curve for MSE lies a bit farther than the blue curve of raw $\hat{q}_x$. We accept those artefacts since the quadratic errors $(f_\theta(x) - \hat{q}_x)^2$ at those ages remain anyway limited since $\hat{q}_x$ is anyway lower. It would be much less optimal for MSE to have a poor fit at higher ages where $\hat{q}_x$ is higher, with potentially a bigger impact on the quadratic errors $(f_\theta(x) - \hat{q}_x)^2$.

Yet, the situation is completely different in the MLE framework (under the gaussian assumption). Parameters are then essentially estimated in order to minimize the sum of quadratic errors, but we account for weights inversely proportional to the variance of the estimator $\hat{q}_x$. As such, not all ages have an equal importance in the optimization procedure. High ages, with a lower population L_x, have a reduced weight in the sum. It is not anymore an imperative to have a model performing well at high ages. Of course, a poor model at high ages still has a tremendous impact on the related quadratic error terms (since $\hat{q}_x$ is already high for high ages), but this impact is anyway reduced due to the lower weight at high ages. As such, it becomes here much more interesting to have an excellent fit in the range of intermediate ages. Due to the weights, it is in that area that a poor fit would have the biggest impact. At high ages, the impact of a poor fit is anyway limited due to the reduced weights (low population). At low ages, the impact of a poor fit is anyway limited since the quadratic errors would be anyway limited for $\hat{q}_x$ quite low.

This is indeed what is observed in fig. 3 (green curve for MLE). The MLE model tries to reach the best fit possible at intermediate ages, even if it comes at the price of a less qualitative fit for high and low ages. In fact, the MLE model benefits from the parameter d (extra degree of freedom for $x \geq 70y.$) in order to reach an excellent fit around age x= 50, 60y. and around x= 80, 90y. The discontinuity is here negative ($d^{MLE} \leq 0$ cf. tab. 7). We have a negative jump that allows to reach an excellent fit for the intermediate ages below 70y. (so around x= 50, 60y.) and for the intermediate ages above 70y. (so around x= 80, 90y.). Indeed, the green curve for MLE is really close to the blue curve of raw $\hat{q}_x$ in those ranges of intermediate ages. It is in that range of intermediate ages that a poor model would have a tremendous impact on the weighted quadratic errors. For the lowest and highest ages below 70y. (so x= 30y. or x=69y.), or for the lowest and highest ages above 70y. (so x= 70y. or x=107y.), the green curve for MLE lies a bit farther than the blue curve of raw $\hat{q}_x$. We accept those artefacts since the weighted quadratic errors at those ages remain anyway quite limited, either since $\hat{q}_x$ is itself limited, either since we have a reduced weight for those ages.

All in all, we see the impact of performing MLE (under the gaussian assumption) or MSE regarding decease probability curves. In the MLE framework, accounting for the weights related to the variance allows us to reach a much better fit of raw $\hat{q}_x$ for the intermediate ages (around x= 70, 90y.) than what is done in MSE. When the MSE framework tries to reach a better fit at high ages, even if the population is quite limited at those high ages, it completely overestimates the mortality in the intermediate range: around x= 70, 90y., the red curve (MSE) is much above the green curve (MLE) in fig. 3.

Life expectancy in the MSE and in the MLE cases

As we have just seen, the MSE model leads to a much higher mortality than the MLE model for the intermediate ages (around x= 70, 90y.). As such, in tab. 8, it is not surprising to recover in all cases a reduced life expectancy for MSE compared to MLE, precisely due to this higher mortality in MSE at intermediate ages.

One could argue that MLE leads to a higher mortality than MSE in the range of very high ages (above x= 90y.). In turn, it should reduce the life expectancy in the MLE framework. This effect is not to neglect and it also participates to the results we observe in tab. 8. The differences between MSE and MLE results are not that huge, since the latter effect at high ages compensates a bit the higher mortality of MSE at intermediate ages. Yet, as we mentioned earlier in section 2.7, the marginal contributions of the successive terms in the life expectancy formula are less and less important for increasing ages. As such, the higher mortality in MLE at high ages is less important regarding the life expectancy than the higher mortality in MSE at intermediate ages. All in all, the life expectancy is reduced in the MSE framework since the effect of the mortality at intermediate ages is predominant.

We can then compare the MSE and MLE life expectancies with the ones extracted from life tables of the moodle website. Both MLE and MSE underestimate a bit the life expectancies from tables. Yet, MLE is systematically closer to the results extracted from tables. This is a new argument to conclude it is important to estimate parameters of a model in a proper log likelihood optimization framework. It highlights again the benefits brought when one accounts for the weights (inversely proportional to the variance of the estimator $\hat{q}_x$) in front of the quadratic errors for the numerical fit.

3 Dynamic mortality modeling

3.1 Introduction to dynamic mortality modeling

In order to catch dynamic effects modifying the mortality over time (such as a reduction of mortality rates and thus an increase of longevity as time passes or cohort effects), prospective models have become a standard. The essential idea is to extrapolate the past trend of mortality rates towards a near future through an adequate predictor. In what follows, we will essentially stick to the simplest but maybe the most usual prospective models: a Lee-Carter predictor in combination either with a log-Poisson model, either with a logit-binomial model. CBD and age-period-cohort models were also described in class but they are not presented in this report for succinctness of the text. This section 3 about prospective mortality modeling is inspired from the article "StMoMo: An R Package for Stochastic Mortality Modeling" [1] that will be referred to as [SMM] in what follows.

3.2 Role of the link function and notations

Let us define q_{xt} and μ_{xt} respectively the instantaneous one-year death probability and instantaneous mortality rate at age x during year t. D_{xt} will be the number of deaths observed at age x during year t in a population counting L_{xt} individuals at age x during year t. E_{xt} will be the exposure to risk at age x, last birthday during year t, according to the definition seen in class. We wish to develop a model for q_{xt}. It is equivalent to develop a model on μ_{xt} given the relation that exists between both quantities. In particular, under the assumption of a piecewise constant mortality rate (constant rate between x and x+1 and between t and t+1 for any age x and year t), we have approximately $\mu_{xt} \approx -ln(1 - q_{xt})$. In practice, we do not have access to the real mortality q_{xt} or μ_{xt}. We only have access to realizations or observations of their unbiased log-likelihood estimators

$$\hat{\mu}_{xt} = \frac{D_{xt}}{E_{xt}} \qquad\qquad \hat{q}_{xt} = \frac{D_{xt}}{L_{xt}}$$

In a Lee-Carter (LC) framework, from observed realizations, we wish to adjust a model of the type

$$g(\mu_{xt}) = g(E(\hat{\mu}_{xt})) = g(E(\frac{D_{xt}}{E_{xt}})) = \eta_{xt} = \alpha_x + \beta_x \kappa_t$$

The quantity η_{xt} represents the predictor of our model, i.e. the systematic component in the model. In what follows, we will use for the predictor a function of the LC-type: $\eta_{xt} = \alpha_x + \beta_x \kappa_t$. The link function g relates per definition the true mortality rate $\mu_{xt} = E(\hat{\mu}_{xt})$ to the predictor η_{xt}. In other words, it relates the random component (observations $\hat{\mu}_{xt}$ whose expectancy are μ_{xt}) to the systematic component.

The question is then how to adjust the systematic model $g(\mu_{xt}) = \eta_{xt}$ for the true mortality to our observations. Several possibilities can be envisioned:

- The original or historical LC model uses a log link function (g=ln) and proposes to regress linearly the logarithm of observations $ln(\hat{\mu}_{xt})$ over the predictor. In that approach, we work under the assumption of homoscedastic mortality rates. In other words, we assume

  $$ln(\hat{\mu}_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

  with ϵ_{xt} a gaussian noise with constant volatility. In practice, this approach might be criticized given several arguments. First of all, there is no particular reason for the noise to be gaussian with constant volatility. In fact, the variance of death rates grows at high ages due to a lower population size at high ages. Besides, the ln function is increasing and concave. By virtue of Jensen's inequality, performing a regression on the logarithm of observations $ln(\hat{\mu}_{xt})$ can only lead to a biased model for the true mortality. The argument is similar to the one presented at the end of section 2.4.

- The log-Poisson framework keeps the same structure for the model but tries to go beyond the previous artefacts. It still elicits a log link function (g=ln) and a predictor of the LC type $\eta_{xt} = \alpha_x + \beta_x \kappa_t$. Yet, it properly relies on a log likelihood maximization framework. Indeed, it assumes D_{xt} to follow a Poisson model $D_{xt} \sim \mathcal{P}(E_{xt}.\mu_{xt})$. Under that assumption, it can be shown that the Poisson likelihood is proportional to

[1] Villegas, A. M., Kaishev, V. K., Millossovich, P. (2018). StMoMo: An R Package for Stochastic Mortality Modeling. Journal of Statistical Software, 84(3), 1–38. https://doi.org/10.18637/jss.v084.i03

the true likelihood. The log likelihood is then maximized. As such, this framework already constitutes an improvement w.r.t. the original LC method.

- The logit-binomial framework tries to go even further. Indeed, even if the Poisson likelihood is proportional to the true likelihood, the real law for D_{xt} must be binomial, not Poisson. In particular, the Poisson framework can lead to more deaths than the size of the population (with very small probability though) whereas it is not possible with the binomial model. As such, the logit-binomial framework assumes $D_{xt} \sim \mathcal{B}(L_{xt}, q_{xt})$. The link function g is not anymore ln but a logit is taken (g=lg). As such, with this framework, we hope to slightly improve the results of the log-Poisson paradigm. Yet, the log-Poisson method is already quite good (true likelihood proportional to Poisson likelihood, and only small probability to have more deaths than the size of the population). In turn, we expect a bigger marginal improvement when going from the original LC model to the log-poisson than when we go from log-Poisson to the logit-binomial. This will be verified in the next section 3.3.

In short, the link function relates the random and the systematic components of the prospective model. In case a Poisson law is used for D_{xt}, it is common to use a log link function. In case a binomial law is used for D_{xt}, it is common to use a logit link function. *"Although a number of link functions would be possible, it is convenient to use the so-called canonical link and pair the Poisson distribution with the log link function and the Binomial distribution with the logit link function"* [SMM]. In what follows, we will try to compare the log-Poisson and logit-binomial frameworks.

3.3 Model fitting

For this section 3.3 and until section 3.6, we will focus on Belgian male data for the years 1960 up to 2010 and for ages x= 1y. to x= 99y. We use the same life tables as before, taken from the moodle website (source mortality.org). Assuming a predictor $\eta_{xt} = \alpha_x + \beta_x \kappa_t$ of the LC-type, we wish now to fit and compare a log-Poisson model and a logit-binomial model.

"The `StMoMo` *package contains the pre-built function* `lc` *to facilitate the definition of Lee-Carter models (with LC-type predictors)"* [SMM]. It is used here with an argument `link = "log"` or `"logit"` to work under the log-Poisson or logit-binomial settings. Then, *"*`StMoMo` *provides the generic function* `fit` *for estimating the parameters of a mortality model"*[SMM]. According to [SMM], a technical manipulation must be performed if one wishes to work under the logit-binomial model. *"In order to match the logit-binomial setting, initial exposures are approximated by transforming the available central exposures. This is accomplished using the utility function* `central2initial` *of package* `StMoMo`*"* [SMM]. As such, this utility function was used in our R-code (cf. appendix C) for the logit-binomial case.

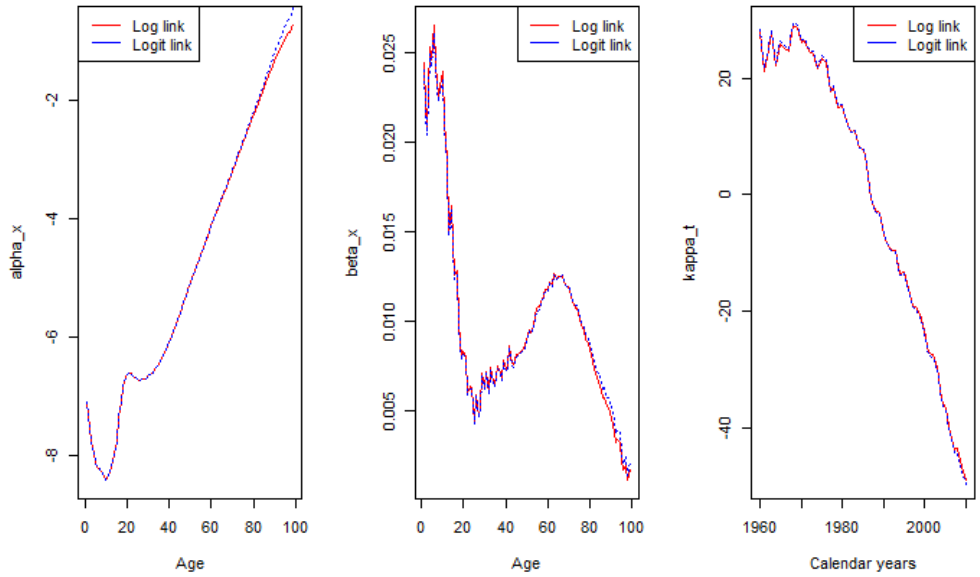

Figure 4: Parameters of the LC-model (both for log and logit links) estimated after the `StMoMo` fit (male data from 1960 to 2010)

The resulting parameters $(\alpha_x, \beta_x, \kappa_t)$ from our fits are plotted in fig. 4 for both cases of log and logit links. The α-parameter *"is a static age function capturing the general shape of mortality by age"* [SMM]. The β-parameter is the marginal reduction of mortality at each age. The κ-parameter captures the evolution of mortality over time. Indeed, as expected, in both settings, we obtain a decreasing κ_t over time and thus a reduction of mortality over time. This will in turn be translated into an increase of longevity over time. The shape of the α-curve is also what we expect from a mortality rate (plotted in log or logit scales). We notice some infant mortality at low ages, a minimum around 15-20y., a bump due to accidents around 21-25y. and then an almost linear increase of the parameter up to high ages.

We observe that the log and logit settings entail very similar parameters, especially for β_x and κ_t. It is not surprising to obtain rather similar models, given what we said in the previous section 3.2. The marginal increment when going from the original LC model to the log-Poisson setting should be greater than the one obtained when going from the log-Poisson to the logit-binomial setting. Log and logit models should be rather close. Regarding the α-parameter, both models are rather similar again. Yet, we notice some edge effects for the log link at high ages. It starts to give lower α values at high ages compared to the logit link. The small marginal improvement we will observe here below when going to the logit link possibly comes from those edge effects at high ages for the α-parameter.

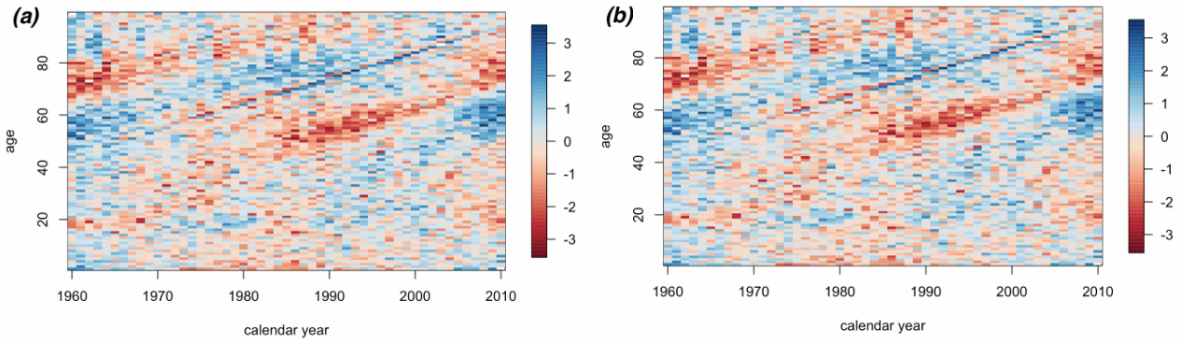

Figure 5: Heat maps of standardized deviance residuals (both for log (a) and logit (b) links) after the `StMoMo` fit (male data from 1960 to 2010)

In order to decide between log and logit links, we must analyze the goodness of our fits. *"The goodness-of-fit of mortality models is typically analyzed by inspecting the residuals of the fitted model. Regular patterns in the residuals indicate the inability of the model to describe all the features of the data appropriately. (...) In* ***`StMoMo`*** *standardized deviance residuals can be obtained with the generic function* ***`residuals`****."* [SMM].

Heat maps of standardized deviance residuals are available both for log and logit links in fig. 5. In both cases, we notice some patterns left in residuals due to cohort effects that were not captured in a LC framework. As mentioned earlier, more advanced models can be envisioned to deal with those cohort effects but they will not be detailed in the present report. The heat maps for log and logit links are also rather similar in terms of order of magnitudes and dispersion. This is again not surprising considering the fact we obtained rather similar parameters $(\alpha_x, \beta_x, \kappa_t)$ for both links. It is a new illustration of the fact that the marginal improvement when going from the log-Poisson to the logit-binomial setting should be rather small (w.r.t. the improvement obtained when going from the original LC model to the log-Poisson).

Finally, *"when evaluating the goodness-of-fit of different models, it has become common in the mortality literature to use information criteria penalizing models with more parameters. Two of these criteria are the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC)"* [SMM]. The latter criteria are defined in [SMM], *"with a lower AIC and BIC being preferable"*. For both log and logit links, the AIC and BIC information was computed using the pre-built R-functions `AIC` and `BIC`. The obtained values are gathered in tab. 10. The lowest values are obtained for the logit link even if the results are rather similar. It is a new illustration of the fact that the marginal improvement when going from the log-Poisson to the logit-binomial setting should exist but it should be rather small (w.r.t. the improvement obtained when going from the original LC model to the log-Poisson). All in all, it is yet an incentive to favour the logit-binomial setting.

Link function	AIC	BIC
Log	46,873.73	48,485.78
Logit	46,578.76	48,190.82

Table 10: AIC and BIC criteria for both log and logit links. Lower AIC and BIC values are an incentive to favour the logit setting.

3.4 Analysis and forecast of the κ_t time series

In prospective mortality modeling, once the model is fitted, it must be extrapolated towards larger time horizons by proper forecast means. In the present situation, it is thus question to extrapolate the only parameter κ_t that depends on time in the modeling of the mortality rate. We must analyze the features of this time series in order to motivate a suitable forecasting model for κ_t.

For most countries, the first order differences of the time series $(\kappa_t)_{t\geq 0}$ are observed to be relatively stationary. We can formally verify this statement in the present case dealing with male data in Belgium from 1960 up to 2010.

In fig. 6, we plotted the first order differences $(\Delta\kappa_t)_t = \kappa_t - \kappa_{t-1}$ as a function of time, both for log and logit links. Again, the plots are very similar for both links given the similarity we observed for both models regarding their parameter κ_t. Yet, what is interesting to notice here is the apparent stationarity of the time series of first order differences. No particular trend can be extracted from those plots in fig. 6. It is a first hint to postulate the stationarity of the time series of first order differences, although it must still be assessed through a proper hypothesis test.

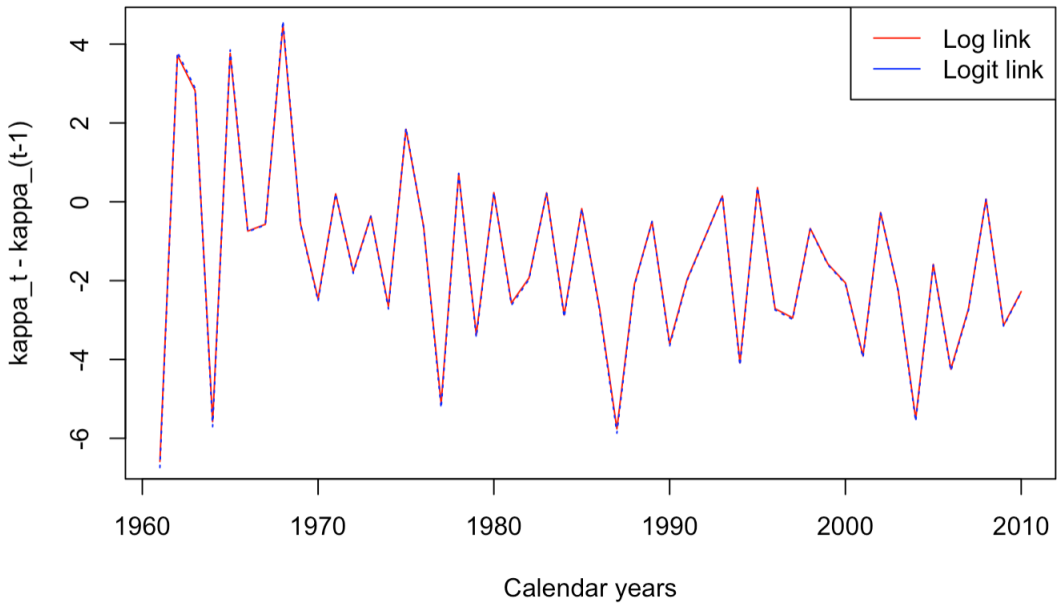

Figure 6: Time series of first order differences $(\Delta\kappa_t)_t = \kappa_t - \kappa_{t-1}$ (both for log and logit links) after the `StMoMo` fit (male data from 1960 to 2010)

We further conducted a Box-Pierce test to assess the null hypothesis of stationarity within the time series of first order differences $(\Delta\kappa_t)_t = \kappa_t - \kappa_{t-1}$. In order to do so, the R-function `Box.test` was applied on those time series of first order differences, with a parameter `lag=20` (to check the absence of auto correlation up to the order `lag`). The resulting p-value of the test in case of a log link is 0.111 whereas the p-value for the logit link amounts to 0.109. Both p-values are above the confidence level $\alpha = 2.5\%$ that we took for the test. The null hypothesis of stationarity is thus not rejected, both for log and logit links.

Given all those arguments, it seems legitimate to postulate one of the simplest ARIMA forecasting model for the time series κ_t:

$$\kappa_{t+1} = \kappa_t + d + \xi_t$$

where the real d is a simple drift and the noise ξ_t is stationary and normally distributed (zero mean and constant volatility σ). This is a simple random walk with a drift.

By default in `StMoMo`, the pre-built function `forecast` extrapolates the period index κ_t using this simple model of random walk with drift [SMM]. Essentially, the drift d and the volatility σ are adjusted *via* their maximum likelihood estimators as seen in class. The function `forecast` was thus applied in both cases of log-Poisson and logit-binomial models. The resulting forecasts of the time series κ_t both for log and logit links are plotted in fig. 7.

A time horizon of 90 years was used for the forecast. This horizon was elicited within the idea that we will have to exploit our extrapolated model in order to assess the life expectancy of a man having 20y. in 2018 (cf. section 3.6). The time horizon of the forecast must thus be sufficient enough in other to catch the essential of the life of such an individual. This will be the only way to catch the essentials of the terms that will appear in the sum used to compute the life expectancy of such a man having 20y. in 2018. If the simulation is conducted until 2100, we anyway exceed the maximal age of x= 99y. considered in the model for such an individual having 20y. in 2018.

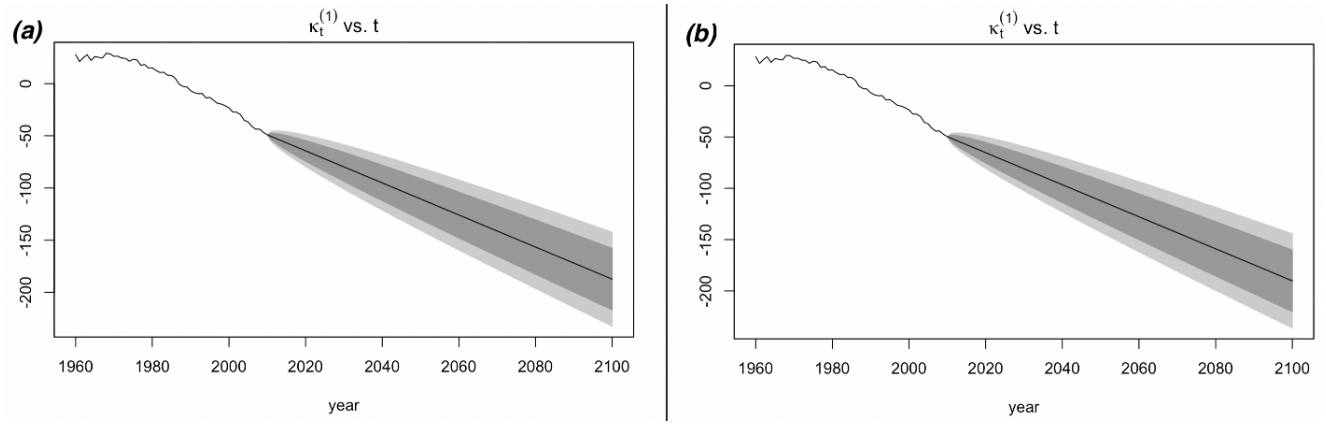

Figure 7: Forecast (horizon 90y.) for the time series κ_t (both for log (a) and logit (b) links) using a simple model of random walk with drift (male data from 1960 to 2010). Dashed lines represent central forecast and dotted lines represent 95% prediction intervals.

3.5 Simulation of mortality curves and discussion

For both log and logit links, for the year 2018, based on the above-mentioned forecasts, we simulated N= 10,000 mortality curves. In order to do so, "the package `StMoMo` provides the pre-built function `simulate` for simulating trajectories from stochastic mortality models" [SMM]. In fig. 8, 15 mortality curves (among the 10,000) were plotted for the year 2018. They give the mortality rate $\mu_{x,2018}$ (in semi-log scale as a function of age) that was obtained in 15 forecast scenarios among the 10,000. The shape of the curves is what we expect from a mortality rate in semi-log scale. We recover an infant mortality part at young ages, a minimum preceding a bump due to accidents around 20-25y. and then an almost linear increase in semi-log scale up to the highest ages.

Across the 10,000 simulated scenarios, it is also possible to compute the mean and the standard deviation for $\mu_{x,2018}$ at any age x= 1y. to x=99y. At any age x in 2018, we have a distribution of $\mu_{x,2018}$ among the 10,000 scenarios, for which we can take the mean and the variance. We can even compute any quantile of this distribution in order to access a confidence interval for $\mu_{x,2018}$ at any age x. For instance, if we take the 0.5% and 99.5% quantiles, we access a 99% confidence interval for $\mu_{x,2018}$ at any age x. Indeed, we know that 99% of the mass of the distribution across our 10,000 forecast scenarios was located within those bounds.

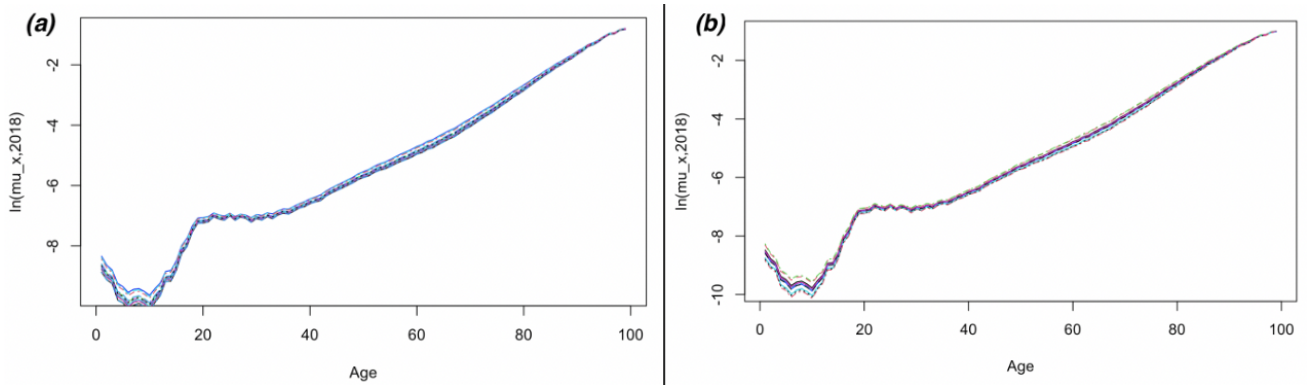

Figure 8: For the year 2018 (both for log (a) and logit (b) links), 15 mortality curves $\mu_{x,2018}$ (in semi-log scale) among the N=10,000 simulated.

The results of those computations, both for log and logit links, are plotted in semi-log scales in fig. 9. We notice that the mean forecast, the standard deviation and the confidence intervals are indeed age-dependent. From fig. 9, it seems that the confidence intervals are reduced at high ages which seems counter-intuitive due to the lower population at high ages. However, we must not forget that our results were plotted in semi-log scales. If we go back to a normal scale for $\mu_{x,2018}$, we indeed verify that the confidence intervals are larger at higher ages, in accordance to the intuition.

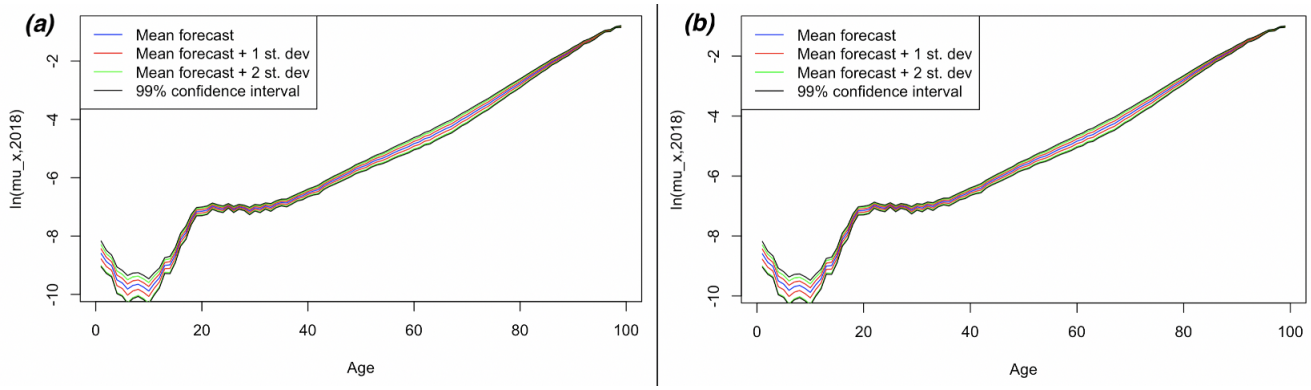

Figure 9: For the year 2018 (both for log (a) and logit (b) links), in semi-log scale, across the N=10,000 simulated scenarios, mean forecast for $\mu_{x,2018}$, standard deviations and 99% confidence interval

Comparison with the observed mortality in 2018

In the life tables provided on the moodle website (source mortality.org), the raw $\hat{\mu}_{xt} = \frac{D_{xt}}{E_{xt}}$ is also given (column "mx"). We wish to compare this observed mortality in 2018 w.r.t. the forecast one in fig. 9.

However, when we plot $\hat{\mu}_{x,2018}$ as a function of age, there are many ages for which this raw mortality lies outside the confidence intervals we had established in fig. 9. This can be noticed in fig. 10 (that concerns the logit link function). The raw or observed mortality $\hat{\mu}_{x,2018}$ corresponds to the red curve. The mean forecast mortality and the 99% confidence interval of previous fig. 9-(b) are plotted again in fig. 10, respectively using the blue curve and the dark curves. Indeed, there are many ages for which the observed mortality $\hat{\mu}_{x,2018}$ (red curve) lies outside the previous confidence interval (dark curves). Yet, something is still missing in the analysis.

The period index κ_t had to be forecast and we had an uncertainty about this forecast (cf. fig. 7). This forecast uncertainty about κ_t translates into a forecast uncertainty about the predictor $\eta_{xt} = \alpha_x + \beta_x \kappa_t$ and thus into a forecast uncertainty about the true $\mu_{xt} = g^{-1}(\eta_{xt})$. The previous confidence intervals of fig. 9 (or the dark curves in fig. 10) only concern this forecast uncertainty about the true component $\mu_{xt} = E(\hat{\mu}_{xt})$.

Besides that, we still have random fluctuations of the observed mortality $\hat{\mu}_{xt}$ around the systematic component $\mu_{xt} = E(\hat{\mu}_{xt})$. We only have access to one realization of observed mortality $\hat{\mu}_{xt}$ but the true mortality μ_{xt} can still lie somewhere around. In other words, we also have a confidence interval for $\mu_{xt} = E(\hat{\mu}_{xt})$ around $\hat{\mu}_{xt}$. This confidence interval does not come anymore from a forecast uncertainty about the systematic component, but from a realization uncertainty for the random component around the systematic component.

This second source of uncertainty (random component $\hat{\mu}_{xt}$ around the systematic component μ_{xt}) is what we try to quantify *via* the green curves in fig. 10. They represent a confidence interval for the true mortality μ_{xt} around the observed one $\hat{\mu}_{xt}$.

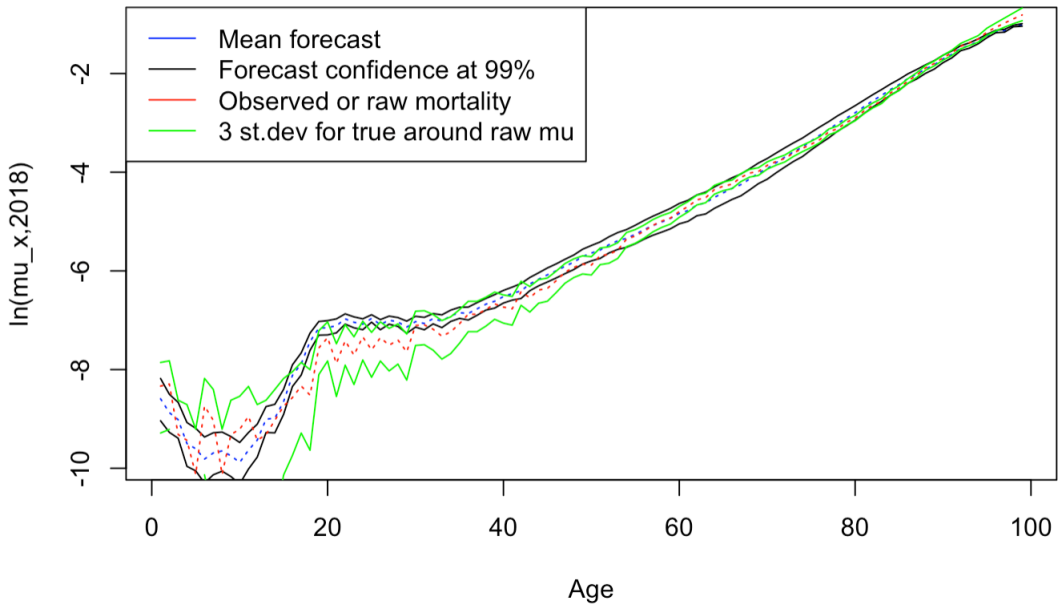

Figure 10: For the year 2018 (for logit link), in semi-log scale, two types of confidence intervals for the true mortality μ_{xt}. A forecast uncertainty at 99% confidence across the N=10,000 simulated scenarios: dark curves around the mean forecast (blue curve). A realization uncertainty: the true mortality μ_{xt} is somewhere around the raw observation $\hat{\mu}_{xt}$ (red curve). We give here a confidence interval with 3 standard deviations for μ_{xt} around $\hat{\mu}_{xt}$ (green curves around red curve).

In summary, we have access to two types of confidence intervals for the true mortality μ_{xt}. The first type is a forecast uncertainty for μ_{xt} around the mean forecast (dark curves around the blue curve in fig. 10). The second type is a realization uncertainty. The true mortality μ_{xt} is somewhere around the observed one $\hat{\mu}_{xt}$ (green curves around the red curve in fig. 10). If both confidence intervals for the true mortality (dark and green curves) have an intersection for most ages, the model appears to be consistent. If so, there is no reason to reject the model since the true mortality predicted from the forecast reasonably explains the true mortality we expect from our observations. This is indeed what is observed in fig. 10 for the logit link. As such, our logit model appears to be consistent. A similar discussion is possible for the log link and the corresponding plot is given in appendix B.

The question was then how to establish the realization confidence intervals (green curves in fig. 10). The following procedure was applied as a first approximation. The life tables of the moodle website (source mortality.org) give access to the raw $\hat{q}_{x,2018}$. The standard deviation of this estimator $\hat{q}_{x,2018}$ was approximated by $\sigma_x = \sqrt{\frac{\hat{q}_{x,2018}(1-\hat{q}_{x,2018})}{L_{x,2018}}}$. From $\hat{q}_{x,2018}$, we computed a confidence interval (at 3 standard deviations) for the one year death probability $q_{x,2018,u/d} = \hat{q}_{x,2018} \pm 3\sigma_x$. Given the approximated link $\mu_{xt} \approx -ln(1-q_{xt})$ between a mortality rate and the death probability (for a piecewise constant rate), we interpreted $\mu_{x,2018,u/d} = -ln(1-q_{x,2018,u/d})$ as the requested realization confidence intervals (or green curves). This reasoning is not at all formal and should be improved in a proper discussion. However, this already gives access to a first order of magnitude for the uncertainty about the true μ_{xt} around the observed one $\hat{\mu}_{xt}$. This realization uncertainty comes in any case in parallel of the forecast uncertainty for the true μ_{xt} around the central forecast.

3.6 Life expectancy under the prospective framework and discussion

Using our previous results, both for log and logit links, we wish now to compute the life expectancy of a man having x= 20, 40 or 60y. in 2018.

For the sake of comparison, we will first consider the life expectancies (at those ages for a man in 2018) that are given in the life tables of the moodle website (column "ex"). The latter are gathered in the first column of tabs. 11 and 12. We will also consider the life expectancies (at those ages for a man in 2018) that are computed under a static framework. The latter are gathered in the second column of tabs. 11 and 12. In order to obtain those static results, we essentially applied the static procedure described in section 2.7 to the smooth q_x that we had obtained for a man in 2018 using the WH smoothing.

A non clever exploitation of the prospective model
There is a first way to exploit our prospective model in order to compute the requested life expectancies. The associated results are given in the third and fourth columns of tab. 11 for the log and logit links. However, we will see that it is not a clever way to proceed since the prospective framework is not interpreted the way it should.

As proposed in the project statement, we can use the average forecast death probabilities in 2018 in order to compute those life expectancies. In other words, from the mean forecast rate $\mu_{x,2018}$ (at any age x in 2018, cf. section 3.5), we can derive a mean forecast death probability $q_{x,2018} \approx 1 - exp(-\mu_{x,2018})$. We simply used the approximated link $\mu_{xt} \approx -ln(1 - q_{xt})$ between a mortality rate and the death probability (for a piecewise constant rate). The life expectancy is then essentially computed as in section 2.7, but with those average death probabilities simulated in 2018 by the prospective model

$$\mathring{e}_x = \sum_{t=1}^{99-x} t.{}_tp_{x,2018}.q_{x+t,2018}$$

$${}_tp_{x,2018} = p_{x,2018}.p_{x+1,2018}...p_{x+t-1,2018} = (1 - q_{x,2018}).(1 - q_{x+1,2018})...(1 - q_{x+t-1,2018}) \qquad (t \geq 1)$$

The summation is here taken up to the maximal age 99y. envisioned in the prospective model.

However, this is not exploiting the full potential of the prospective model. A person having x= 20y. in 2018 will suffer $q_{x=20,2018}, q_{x=21,2019}, q_{x=22,2020}$. Those probabilities should be the ones to consider in the summation formula for the life expectancy. In our non clever exploitation of the prospective model, by only considering the mean $q_{x,2018}$ in 2018 at various ages x, we completely forgot to follow over time the cohort to which this individual belongs. The $q_{x,2018}$ at various ages x are in fact relative to different cohorts, not only to the cohort of the individual in question. In short, we used the prospective model as we would have run a static computation (only exploiting the mortality in 2018 whatever the cohort) which is not clever at all.

As noticed in tab. 11, the obtained results (from this non clever usage) completely underestimate the life expectancies extracted from life tables or obtained from a static computation. In a dynamic framework, we would expect to catch a decrease of mortality over time and thus an increase of life expectancies compared to a static framework. This is not what is observed in tab. 11. This is because we made a non clever usage of the prospective framework. We exploited the dynamic paradigm as we would have run a static computation.

	Life tables	Static framework	Log link non clever usage	Logit link non clever usage
$\mathring{e}_{20}$	59.74	59.24843	57.95118	57.7769
$\mathring{e}_{40}$	40.45	39.9524	39.01746	38.9485
$\mathring{e}_{60}$	22.38	21.88842	21.26211	21.30082

Table 11: For a man having x= 20, 40, 60y. in 2018, life expectancies extracted from life tables, computed under a static framework or exploiting the propsective model in a non clever manner for log and logit links (static usage of the prospective model)

The clever exploitation of the prospective model
To resolve those artefacts of the non clever exploitation, we must follow over time the cohort to which the individual belongs. Men having x= 20, 40 or 60y. in 2018 respectively belong to the cohorts 1998, 1978 and 1958. The `StMoMo` pre-built function `extractCohort` allows us to access the mean mortality rates μ_{xt} (average across the N= 10,000

simulated scenarios) that the cohort will have to face over time. In other words, for the cohort 1998, we can extract $\mu_{x=20,2018}, \mu_{x=21,2019}, \mu_{x=22,2020}, ...$ and similarly for the two other cohorts. It is then easy to compute the death probabilities $q_{xt} \approx 1 - exp(-\mu_{xt})$ that the cohort will face over time. We simply used the approximated link $\mu_{xt} \approx -ln(1 - q_{xt})$ between a mortality rate and the death probability (for a piecewise constant rate). The life expectancy is then essentially computed as in section 2.7, but with those average death probabilities (simulated by the prospective model) that the cohort will face over time. For instance, for the 1998 cohort,

$$\mathring{e}_{20} = \sum_{t=1}^{99-20} t._tp_{20}.q_{20+t,2018+t}$$

$$_tp_{20} = p_{20,2018}.p_{21,2019}...p_{20+t-1,2018+t-1} = (1 - q_{20,2018}).(1 - q_{21,2019})...(1 - q_{20+t-1,2018+t-1}) \qquad (t \geq 1)$$

The summation is here taken up to the maximal age 99y. envisioned in the prospective model. We can now understand why the horizon for the forecast/simulation had to be taken large enough (cf. horizon 90y. we spoke about in section 3.4). The cohort 1998 will have x= 99y. (maximal age in the prospective model) around 2097.

As noticed in tab. 12 (third and fourth columns), the obtained results from this clever usage of the prospective model are much more realistic. In particular, this clever usage of the prospective model catches the decrease of mortality over time and thus the increase of life expectancies compared to a static framework.

	Life tables	Static framework	Log link clever usage	Logit link clever usage
$\mathring{e}_{20}$	59.74	59.24843	61.47985	60.55755
$\mathring{e}_{40}$	40.45	39.9524	41.63907	41.22625
$\mathring{e}_{60}$	22.38	21.88842	22.42917	22.35194

Table 12: For a man having x= 20, 40, 60y. in 2018, life expectancies extracted from life tables, computed under a static framework or exploiting the propsective model in a clever manner for log and logit links (full exploitation of the prospective model)

3.7 Temporary life annuity under the prospective framework and discussion

For this section 3.7, we will change the dataset, considering Belgian male data between 1960 and 2018. Similarly to what was done in section 3.3, we tried again to fit a prospective model using a predictor of the LC-type $\eta_{xt} = \alpha_x + \beta_x \kappa_t$, either under a log-Poisson setting, either under a logit-binomial setting. Parameters $(\alpha_x, \beta_x, \kappa_t)$ were adjusted the same way and the discussion about the goodness-of-fit is very similar. For both log and logit links, the AIC and BIC information was computed, giving the results gathered in tab. 13. Without a surprise (considering the same arguments as in section 3.3), the lowest AIC and BIC are obtained for the logit link, which was an incentive to elicit the logit-binomial setting for what follows.

Link function	AIC	BIC
Log	55,183.63	56,885.07
Logit	54,808.5	56,509.94

Table 13: AIC and BIC criteria for both log and logit links. Lower AIC and BIC values are an incentive to favour the logit setting.

Exploiting thus the logit-binomial prospective model, we wish now to price (net single premium) a temporary life annuity (duration n= 20y., annual payment of C=1) for a man having x= 60y. in 2018, under the assumption that $i = 0\%$. In order to do so, the model was extrapolated and simulated the same way we had done in sections 3.4 and 3.5 (N= 10,000 scenarios simulated). The simulation was conducted on a horizon of 50y. which is indeed greater than the duration of the product we wish to price.

A person having x= 60y. in 2018 belongs to the cohort 1958. Each scenario (j) simulated (j= 1,2,... 10,000) delivers a forecast for the mortality rate that this cohort 1958 will have to face:

$$\mu^{(j)}_{x=60,2018}; \quad \mu^{(j)}_{x=61,2019}; \quad \dots \quad \mu^{(j)}_{x=80,2038}$$

where the notation (j) recalls that such a forecast of the mortality rate can be found for any scenario j= 1,2,... 10,000. Using the approximated link $\mu_{xt} \approx -ln(1-q_{xt})$ between a mortality rate and the death probability (for a piecewise constant rate), we can deduce, for any scenario j= 1,2,... 10,000, a forecast for the decease probability that the cohort 1958 will have to face:

$$q^{(j)}_{x=60,2018} \approx 1-exp(-\mu^{(j)}_{x=60,2018}); \quad q^{(j)}_{x=61,2019} \approx 1-exp(-\mu^{(j)}_{x=61,2019}); \quad \dots \quad q^{(j)}_{x=80,2038} \approx 1-exp(-\mu^{(j)}_{x=80,2038})$$

For each scenario j= 1,2,... 10,000, this forecast of death probabilities for the cohort 1958 is used in order to price the temporary life annuity. Under the assumption that $i = 0\%$, the actualisation factor v reduces to the unity such that:

$${}_{20}\ddot{a}^{(j)}_{60} = \sum_{k=0}^{19} {}_kp^{(j)}_{60}\, v^k = \sum_{k=0}^{19} {}_kp^{(j)}_{60}$$

where the notation (j) recalls that such a price can be found for any scenario j= 1,2,... 10,000. Of course ${}_0p^{(j)}_{60} = 1$ for any scenario j= 1,2,... 10,000. The remaining survival probabilities ${}_kp^{(j)}_{60}$ (if $k \geq 1$) can be computed as follows:

$${}_kp^{(j)}_{60} = p^{(j)}_{60,2018} \cdot p^{(j)}_{61,2019} \cdots p^{(j)}_{60+k-1,2018+k-1} = (1-q^{(j)}_{60,2018}).(1-q^{(j)}_{61,2019}) \cdots (1-q^{(j)}_{60+k-1,2018+k-1})$$

As such, we obtain a distribution of the price for the temporary life annuity across the N= 10,000 scenarios, distribution for which we can compute the mean, the standard deviation or any quantile. In particular, we can consider the quantiles 2.5% and 97.5% of the distribution in order to access a 95% confidence interval for the price. Indeed, we know that 95% of the mass of the distribution of the price we had simulated was located between those bounds. The results of those computations (for the elicited logit link model) are to be found in tab. 14.

For the sake of comparison, we also added in the first column of tab. 14 the net single premium (for the same temporary life annuity) that we can compute under a static framework (with Belgian male data of 2018). This static result was essentially computed the same way as for the life annuity in section 2.3, except that x= 60y., that C=1 and that this new product as a finite duration of n= 20y. (the sum must only be taken up to n-1). The smooth q_x for male data in 2018 obtained by WH smoothing were used.

It is not surprising to find a static premium below the mean premium obtained in the prospective case. In fact, here, the static premium lies even outside the 95% confidence interval obtained in the prospective computations. For a contract guaranteeing life benefits, the reduction of mortality over time accounted for in the prospective framework should naturally increase the requested premiums compared to the static case. This is what is observed here.

Static ${}_{20}\ddot{a}_{60}$	Mean ${}_{20}\ddot{a}_{60}$ (prospective logit)	St. dev. for ${}_{20}\ddot{a}_{60}$ (prospective logit)	Quantile 2.5% for ${}_{20}\ddot{a}_{60}$ (prospective logit)	Quantile 97.5% for ${}_{20}\ddot{a}_{60}$ (prospective logit)
17.32927	17.75526	0.1244088	17.5035	17.98736

Table 14: For a man having x= 60y. in 2018, net single premium of a temporary life annuity (duration n= 20y., $i = 0\%$, annual payment of C=1). The first column is the result obtained in a static framework. The other columns give the average price obtained in a prospective framework (logit link), the standard deviation and some quantiles of this price.

A Appendix: Results for WH smoothing

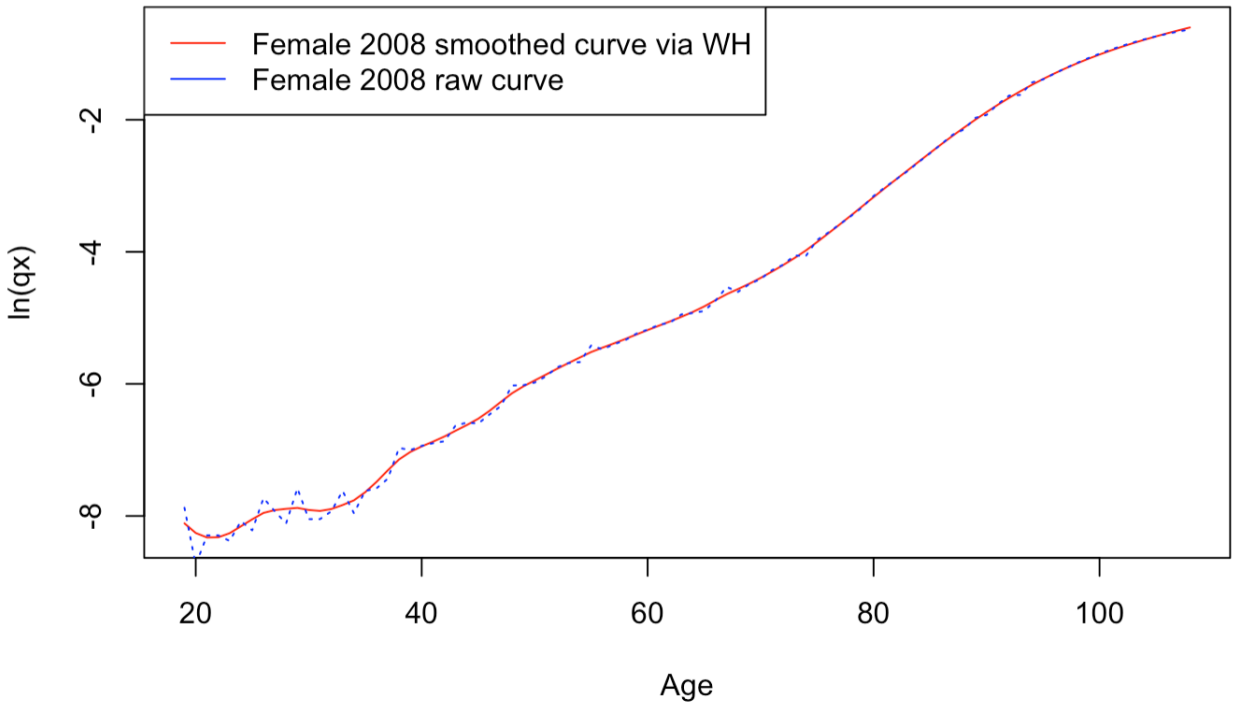

Figure 11: One-year death probability: raw $\hat{q}_x$ and smoothed curve $q_x^s(h)$ obtained by WH smoothing (female 2008)

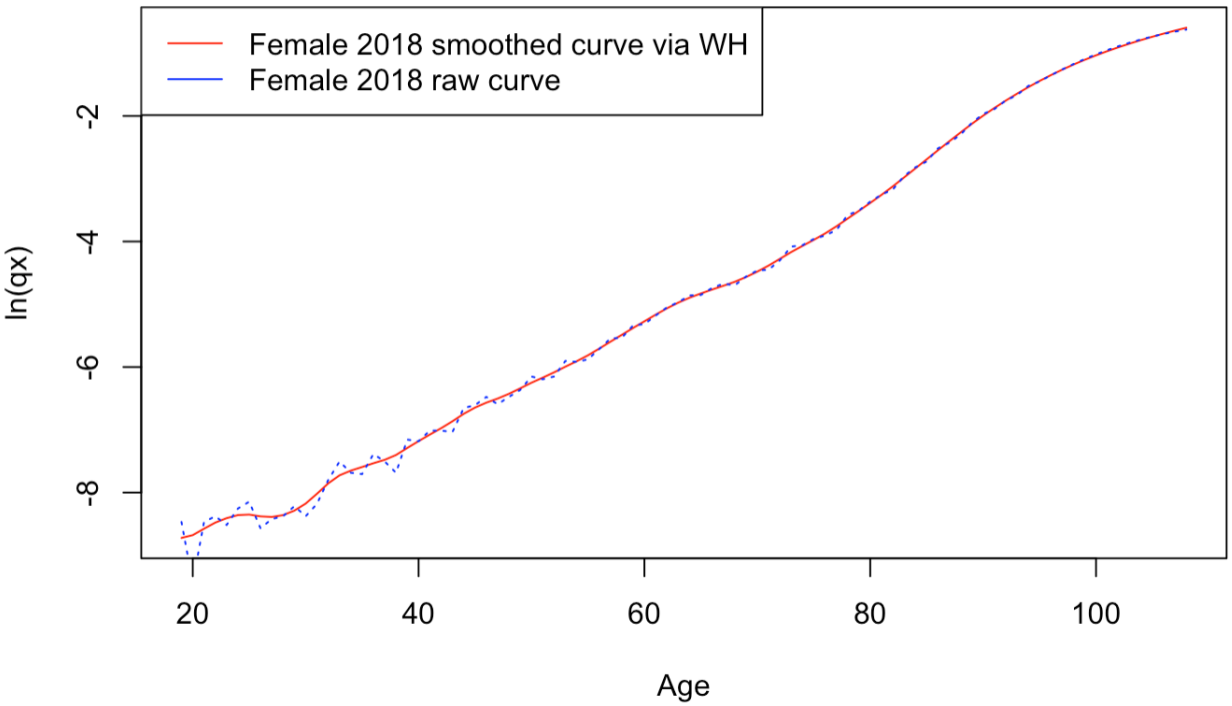

Figure 12: One-year death probability: raw $\hat{q}_x$ and smoothed curve $q_x^s(h)$ obtained by WH smoothing (female 2018)

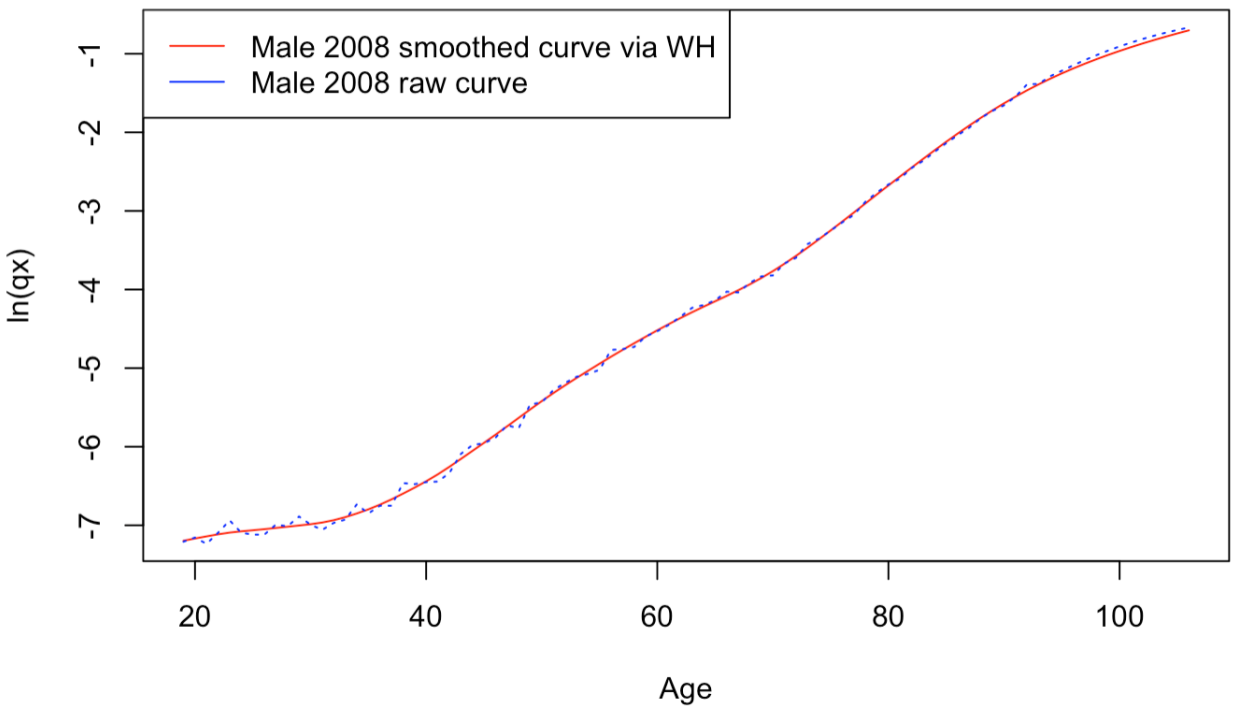

Figure 13: One-year death probability: raw $\hat{q}_x$ and smoothed curve $q_x^s(h)$ obtained by WH smoothing (male 2008)

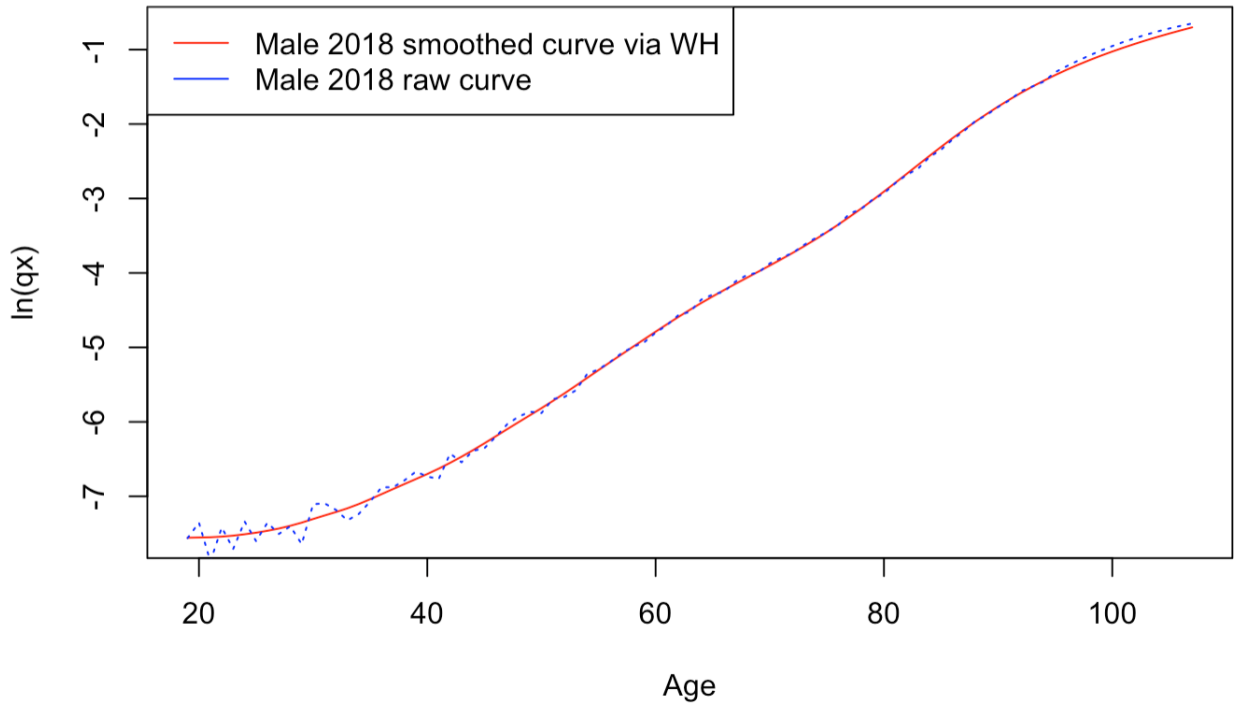

Figure 14: One-year death probability: raw $\hat{q}_x$ and smoothed curve $q_x^s(h)$ obtained by WH smoothing (male 2018)

B Appendix: Confidence intervals for the mortality in prospective modeling

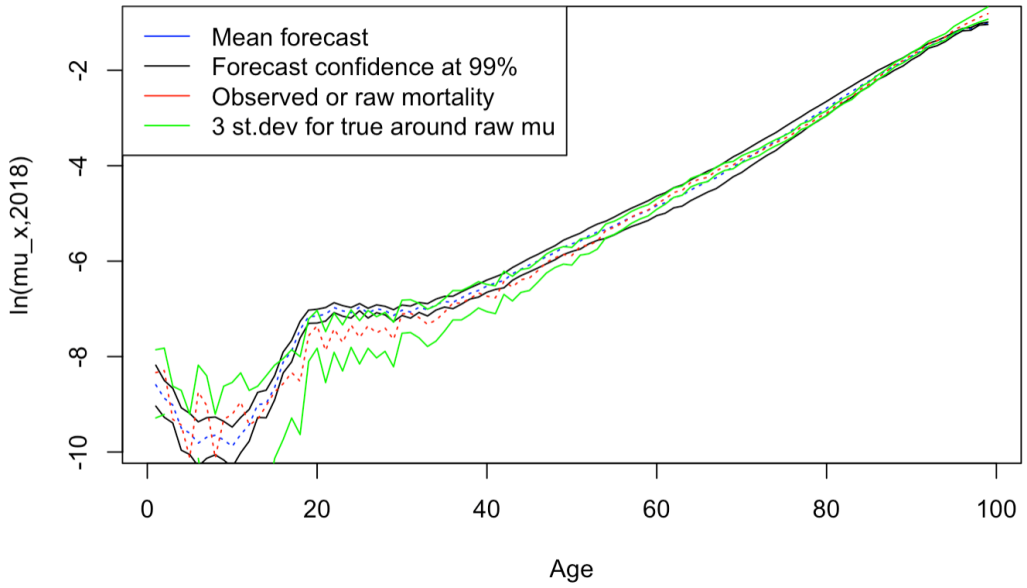

Figure 15: For the year 2018 (for logit link), in semi-log scale, two types of confidence intervals for the true mortality μ_{xt}. A forecast uncertainty at 99% confidence across the N=10,000 simulated scenarios: dark curves around the mean forecast (blue curve). A realization uncertainty: the true mortality μ_{xt} is somewhere around the raw observation $\hat{\mu}_{xt}$ (red curve). We give here a confidence interval with 3 standard deviations for μ_{xt} around $\hat{\mu}_{xt}$ (green curves around red curve).

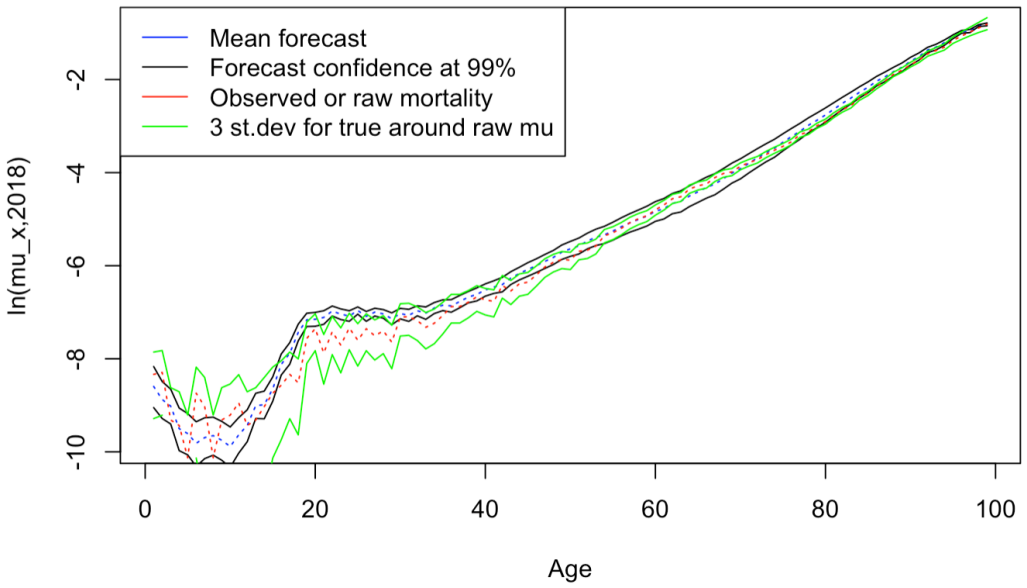

Figure 16: For the year 2018 (for log link), in semi-log scale, two types of confidence intervals for the true mortality μ_{xt}. A forecast uncertainty at 99% confidence across the N=10,000 simulated scenarios: dark curves around the mean forecast (blue curve). A realization uncertainty: the true mortality μ_{xt} is somewhere around the raw observation $\hat{\mu}_{xt}$ (red curve). We give here a confidence interval with 3 standard deviations for μ_{xt} around $\hat{\mu}_{xt}$ (green curves around red curve).