Non mean reverting affine processes for stochastic mortality

Elisa Luciano, University of Turin and ICER
Elena Vigna, University of Turin

The mathematical tools: some basic ideas

The default time is often modelled as a "doubly stochastic stopping time".

The definition of doubly stochastic stopping times needs some preliminaries:

1) A counting process N_t is said to admit the stochastic intensity λ (where λ is a nonnegative predictable process s.t. $E(\int_0^t \lambda_u du) < \infty$) if $M_t = N_t - \int_0^t \lambda(u)du$ is a martingale.

2) If a counting process N_t admits the intensity λ, then

$$E(N_{t+\Delta t} - N_t|\mathcal{F}_t) = \lambda_t \Delta t + o(\Delta t)$$

In other words, the process λ gives information about the average number of jumps of the process under observation in a small period of future time.

3) A counting process with stochastic intensity is *doubly stochastic driven by the subfiltration* $\mathcal{G}_t \subset \mathcal{F}_t$ if, conditionally on the path of λ until s, the process $N_s - N_t$ has Poisson distribution with parameter $\int_t^s \lambda_u du$.

Stopping time

The stopping time of a doubly stochastic process is the analogous of the first jump time of a Poisson process, where the intensity is a stochastic process.

If τ is the first jump time of a Poisson process with parameter λ, then τ has exponential distribution and

$$P(\tau > t) = e^{-\lambda t}$$

Similarly, an important result for doubly stochastic stopping times is that if τ is doubly stochastic with intensity λ, then:

$$P(\tau > t|\mathcal{F}_s) = E\left[e^{-\int_s^t \lambda(u)du}|\mathcal{F}_s\right] \qquad (\star)$$

If the stopping time τ is chosen to be the random time of death of an individual aged x at time s, T_x, then the probability $(\star)$ is the survival probability ${}_tp_x$.

The affine framework

The choice of the intensity process is crucial for the solution of ($\star$).

Classical results from the credit risk literature show that if the process chosen for the intensity is of the affine class, then the expectation ($\star$) turns out to be tractable.

A process λ_t is affine if it is a jump-diffusion process, i.e. if it can be described by the SDE :

$$d\lambda_t = \mu(\lambda_t)dt + \sigma(\lambda_t)dB_t + dJ_t$$

where J is a pure jump process and where the drift $\mu(\lambda_t)$, the covariance matrix $\sigma(\lambda_t)\sigma(\lambda_t)'$ and the jump measure associated with J have affine dependence on λ_t.

Examples of affine processes in finance: Vasicek, CIR.

IMPORTANT RESULT

If λ is affine:

$$E\left[e^{\int_t^T -\lambda(u)du}|\mathcal{G}_t\right] = e^{\alpha(T-t)+\beta(T-t)\lambda(t)} \qquad (\star\star)$$

where the coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalized Riccati ODEs, that can be solved at least numerically and in some cases analytically (Duffie Pan Singleton, 2000).

The actuarial application

We consider an individual aged x and model her random future lifetime T_x as a doubly stochastic stopping time with intensity λ_x.

According to ($\star$) the survival probability is:

$$S_x(t) = P(T_x > t|\mathcal{G}_0) = E\left[e^{-\int_0^t \lambda_x(u)du}|\mathcal{G}_0\right]$$

Previous (recent) literature on this: Biffis (2004), Dahl (2004), Shrager (2004).

The crucial point now becomes: how do we choose the process λ so that to apply the useful equation ($\star\star$)?

First application: mean reverting processes

In the credit risk literature, mean reverting processes work quite well to describe the intensity of default (Duffie and Singleton, 2003):

1. CIR process:

$$\lambda_x(t) = k(\gamma - \lambda_x(t))dt + \sigma\sqrt{\lambda_x(t)}dW(t)$$

2. mean reverting with jumps (m.r.j.):

$$d\lambda_x(t) = k(\gamma - \lambda_x(t))dt + dJ(t)$$

3. VASICEK process:

$$d\lambda_x(t) = k(\gamma - \lambda_x(t))dt + \sigma dW(t)$$

with $k > 0, \gamma > 0, \sigma > 0, W(t)$ standard Brownian motion, $J(t)$ compound Poisson process with intensity l and jumps exponentially distributed with expected value μ.

Survival function

These processes are affine and we can apply ($\star\star$) and express the survival function in closed form (these are standard results):

$$P(T_x > t|\mathcal{F}_0) = S_x(t) = e^{\alpha(t)+\beta(t)\lambda_x(0)}$$

where:

1. CIR process:

$$\alpha(t) = -\frac{2k\gamma}{\sigma^2}\ln\left(\frac{c+de^{bt}}{b}\right) + \frac{k\gamma}{c}t \qquad \beta(t) = \frac{1-e^{bt}}{c+de^{bt}}$$

$$b = -\sqrt{k^2+2\sigma^2} \quad c = 0.5(b-k) \quad d = 0.5(b+k)$$

2. mean reverting with jumps process:

$$\alpha(t) = -\gamma(t+\beta(t)) - l\frac{\mu t - \ln(1-\mu\beta(t))}{\mu+k}$$

$$\beta(t) = \frac{e^{-kt}-1}{k}$$

3. Vasicek process:

$$\alpha(t) = -\frac{(\beta(t)+t)(k^2\gamma - \frac{\sigma^2}{2})}{k^2} - \frac{\sigma^2\beta(t)^2}{4k}$$

$$\beta(t) = \frac{e^{-kt}-1}{k}$$

The calibration to the UK population

Since $\lambda_x(t)$ is the intensity of mortality at age $x + t$ of an individual aged x at time 0, we choose a generation mortality table and not a contemporaries one.

Two observed generation tables (1880 and 1900, HMD data) and two projected mortality tables (1935, 1945).

Assumptions for the calibration:

- initial age $x = 65$, for both males and females
- jump size $\mu < 0$ (to capture improvements in mortality rates)
- $\lambda_{65}(0) = -\ln(p_{65})$

We minimize the sum of the squared differences between the survival probabilities of the relevant table and the ones implied by the model, and compute the calibration error.

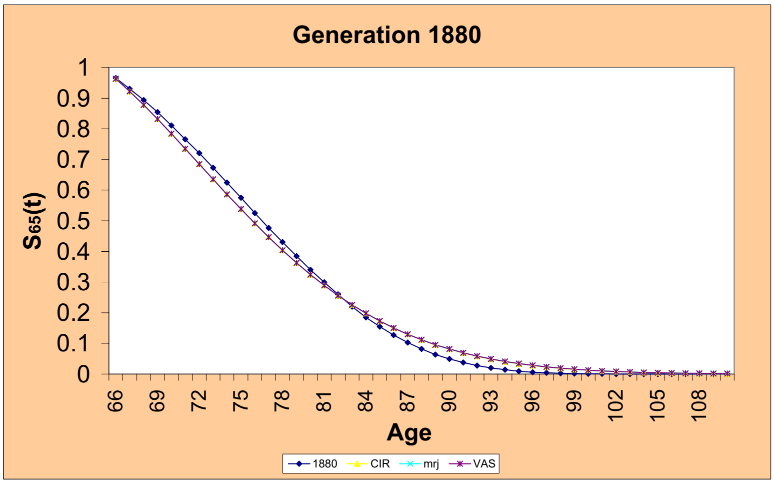
Results from the calibration: value of the optimal parameters

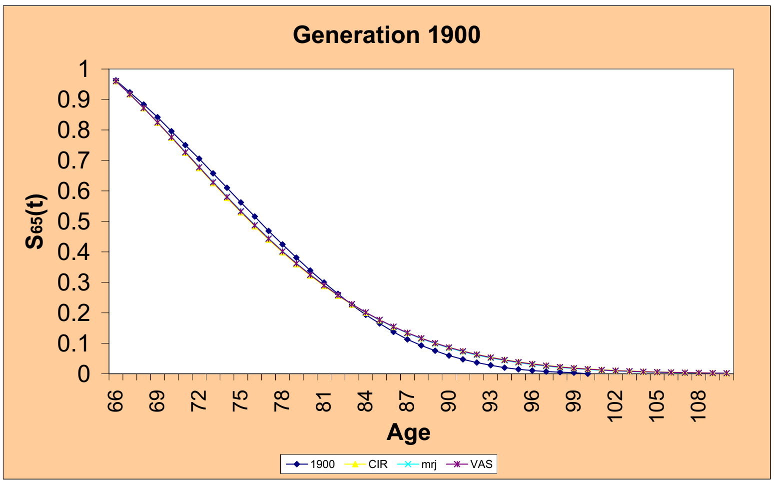
TABLE 1

	1880	1900	1935	1945
$\lambda_{65}(0)$	0.03515	0.03797	0.01145	0.00885
CIR-error	0.02182	0.01662	0.40945	0.20552
CIR-k	0.00448	0.01365	0.06494	0.0078
CIR-σ	0.00103	0.00298	0.00005	0
CIR-γ	1.24656	0.4301	0.07552	0.41711
mrj-error	0.02236	0.01327	0.15816	0.1965
mrj-k	0.00571	0.00392	0.005	0.00465
mrj-μ	-0.00246	-0.00227	-0.00249	-0.00492
mrj-l	0.00247	0.00234	0.00249	0.0099
mrj-γ	0.99382	1.31818	0.64908	0.67935
VAS-error	0.02247	0.01473	0.16191	0.1982
VAS-σ	0.00046	0.00048	0.00002	0.00002
VAS-k	0.00591	0.00835	0.00604	0.00526
VAS-γ	0.96029	0.65393	0.53278	0.59302

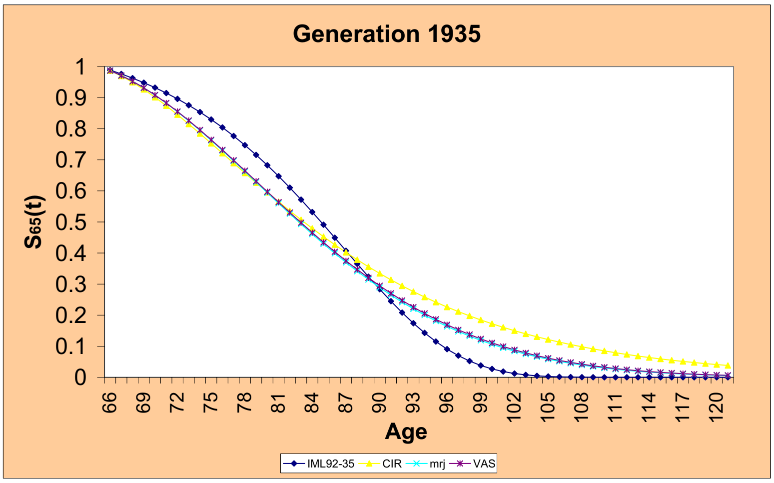
Result: the error more than decuplicates when passing from old observed tables to the projected tables for younger generations.

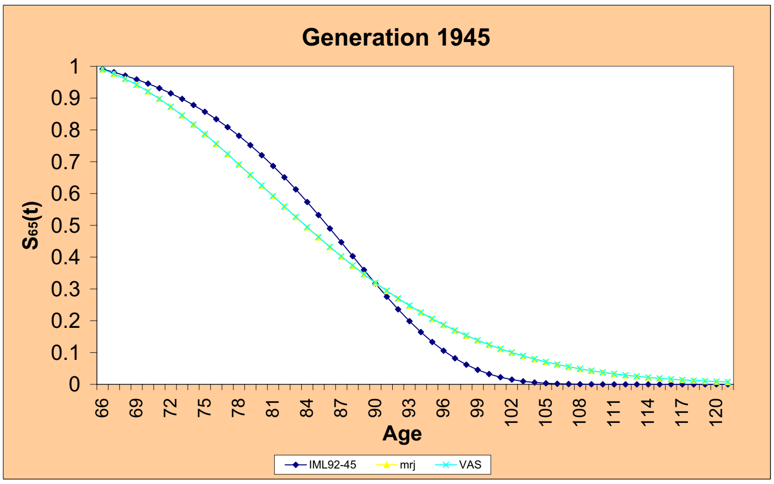
Results from the calibration: survival functions for the older generations





Results from the calibration: survival functions for the younger generations





Evidence:

- fit more satisfactory for the old generations
- for the younger generations, the rectangularization phenomenon is not captured
- the expansion feature is also not captured
- the survival probability at old ages is much higher and at lower ages much lower than in the tables
- the survival probability at very old ages (like 130-140) is positive

Conclusion: in the presence of high rectangularization phenomenon – which is an expected feature in the future generation tables – the intensity of mortality cannot be properly described by the three proposed processes.

QUESTION: is the bad fit due to the common feature of mean reversion of the three models (see also Cairns, Blake and Dowd, 2004)?

OBSERVATION: the force of mortality shows no mean reversion, but rather an exponential increase

SIMPLE IDEA: why not dropping the mean reversion term and calibrate a process whose non stochastic part increases exponentially?

Second application: non mean reverting processes

We propose four models:

1. Ornstein Uhlenbeck process without jumps (OU):

$$d\lambda(t) = a\lambda(t)dt + \sigma dW(t)$$

2. Ornstein Uhlenbeck process with jumps (OUj):

$$d\lambda(t) = a\ \lambda(t)dt + \sigma dW(t) + dJ(t)$$

3. Feller process without jumps (FEL):

$$d\lambda(t) = a\lambda(t)dt + \sigma\sqrt{\lambda(t)}dW(t)$$

4. Feller process with jumps (FELj):

$$d\lambda(t) = a\lambda(t)dt + \sigma\sqrt{\lambda(t)}dW(t) + dJ(t)$$

with $a > 0$ and $\sigma \geq 0$ and J pure compound Poisson jump process, with Poisson arrival times of intensity $l > 0$ and exponentially distributed jump sizes with mean $\mu < 0$

Survival function

The survival function in closed form is:

$$P(T_x > t|\mathcal{F}_0) = S_x(t) = e^{\alpha(t)+\beta(t)\lambda_x(0)}$$

where:

1. OU process:

$$\alpha(t) = \frac{\sigma^2}{2a^2}t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3}$$

$$\beta(t) = \frac{1}{a}\left(1 - e^{at}\right)$$

   Problem: the intensity λ can become negative. The probability of λ becoming negative is:

$$P(\lambda(t) \leq 0) = P\left(N \leq -\frac{\lambda(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}\right)$$

This probability in the applications is negligible.

2. OUj process:

$$\alpha(t) = (\frac{\sigma^2}{2a^2} + \frac{la}{a-\mu})t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} +$$

$$+\frac{l}{a-\mu}\ln(1 - \frac{\mu}{a} + \frac{\mu}{a}e^{at})$$

$$\beta(t) = \frac{1}{a}\left(1 - e^{at}\right)$$

3. FEL process:

$$\alpha(t) = 0 \qquad \beta(t) = \frac{1-e^{bt}}{c+de^{bt}}$$

4. FELj process:

$$\alpha(t) = \frac{l\mu}{c-\mu}t - \frac{l\mu(c+d)}{b(d+\mu)(c-\mu)}\cdot$$

$$\cdot[\ln(\mu - c - (d+\mu)e^{bt}) - \ln(-c-d)]$$

$$\beta(t) = \frac{1-e^{bt}}{c+de^{bt}}$$

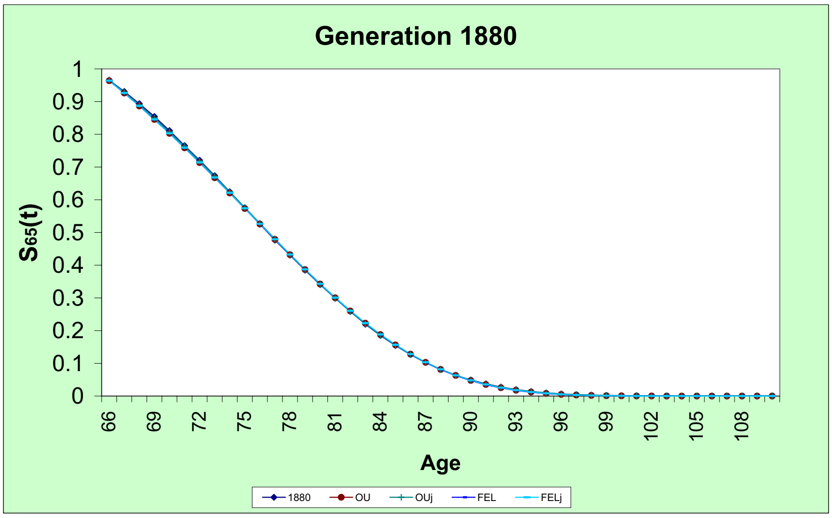
Calibration of non mean reverting processes

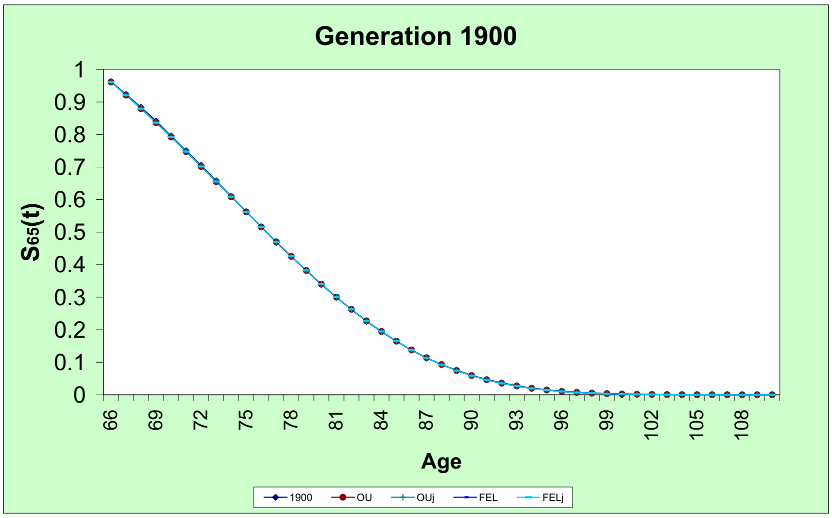
TABLE 2

	1880	1900	1935	1945
$\lambda_{65}(0)$	0.03515	0.03797	0.01145	0.00885
OU-error	0.00043	0.00012	0.00085	0.00027
OU-a	0.0861	0.07949	0.09856	0.10859
OU-σ	0.00183	0.00341	0.0001	0.00048
OUj-error	0.0001	0.00004	0.00002	0.00016
OUj-a	0.09101	0.08192	0.10014	0.10865
OUj-σ	0.00377	0.00414	0.0001	0.00011
OUj-l	0.00173	0.00088	0.00105	0.00036
OUj-μ	-0.00003	-0.00003	-0.00003	-0.00003
FEL-error	0.00044	0.00012	0.00084	0.00027
FEL-a	0.08553	0.07896	0.09867	0.10811
FEL-σ	0.00431	0.01348	0.00005	0.0001
FELj-error	0.00043	0.00012	0.00053	0.00027
FELj-a	0.0858	0.07897	0.10164	0.10811
FELj-σ	0.00735	0.01349	0	0.00001
FELj-l	0.001	0.001	0.1856	0.001
FELj-μ	-0.0001	-0.0001	-0.00034	-0.0001

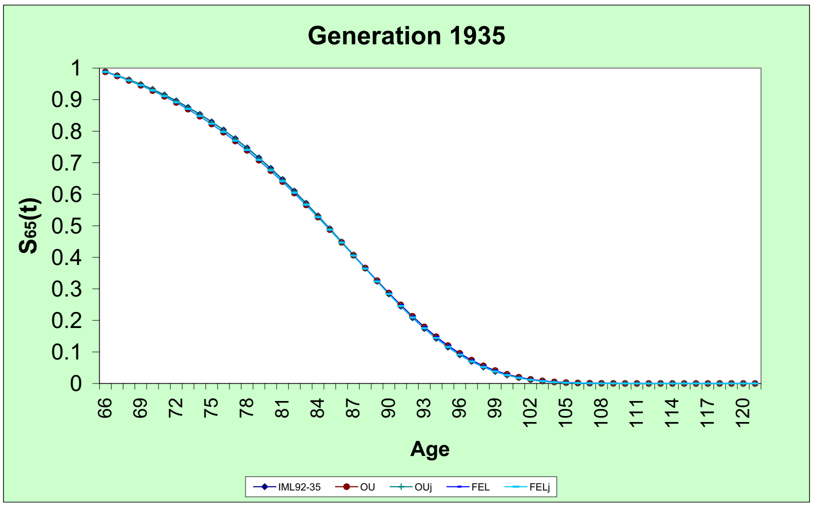
- The calibration errors are very small.
- The OUj model dominates the others. The models with jumps perform better than the corresponding models without.
- The value of σ is very low.

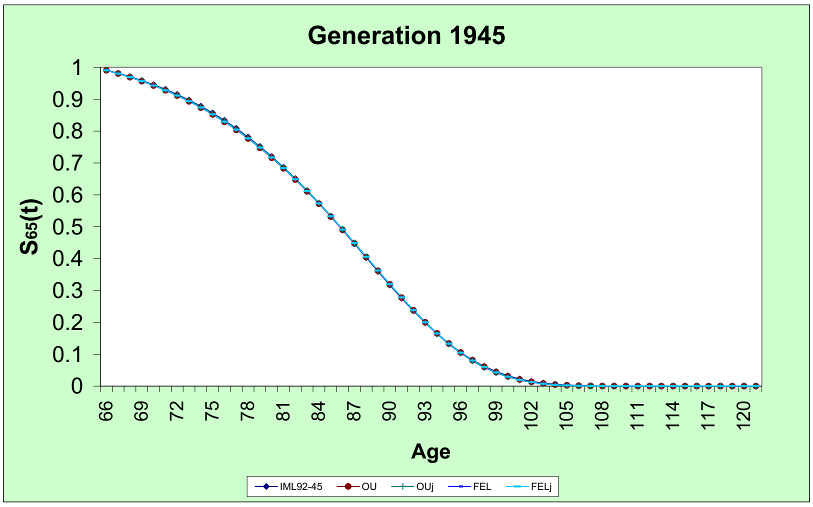
Results from the calibration: survival functions for the older generations





Survival functions for the younger generations

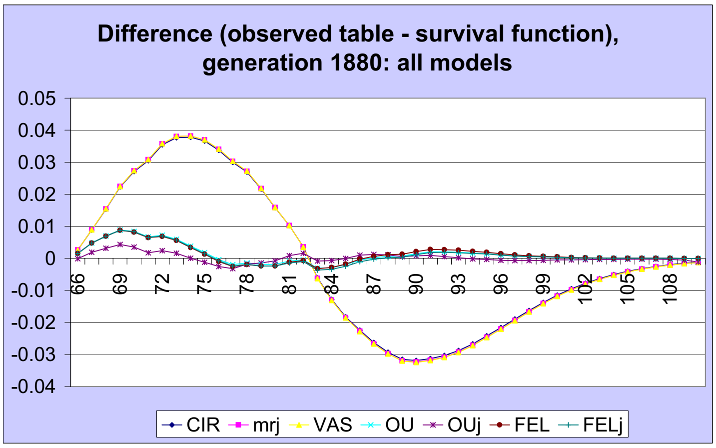


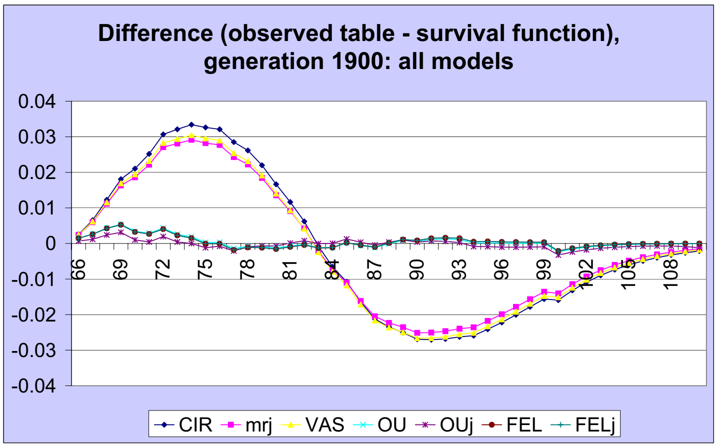


The fit is remarkable, in all cases.

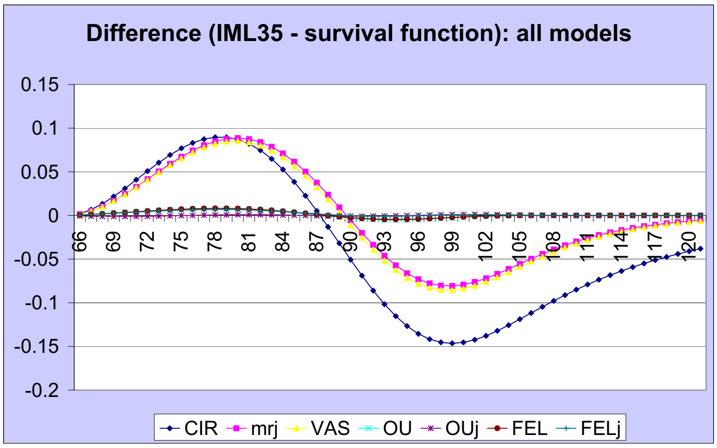
Differences between $S_x(t)$ and ${}_tp_x$

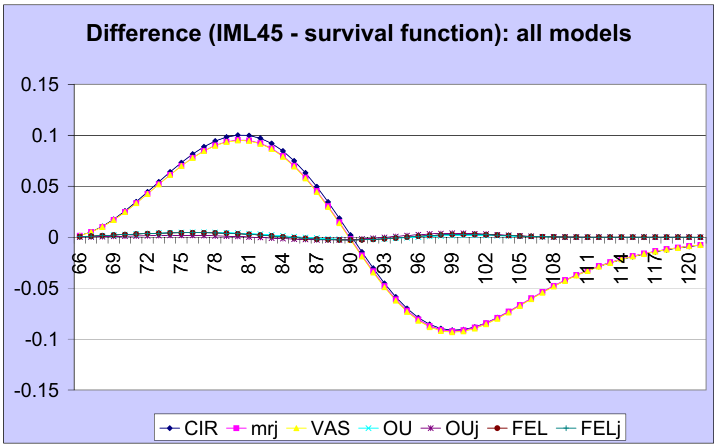
To have a better idea of the goodness of the fit, we plot the differences between the survival function that has to be calibrated and the one implied by the seven models.





Differences between $S_x(t)$ and $_tp_x$

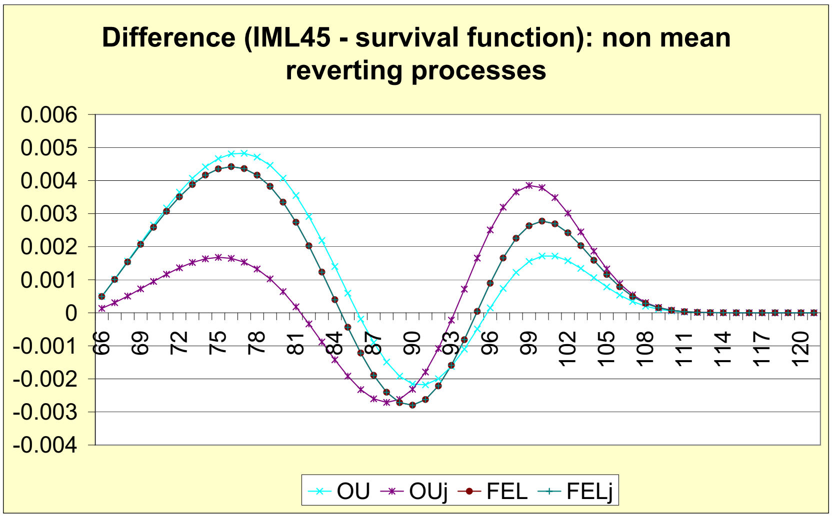




$\Rightarrow$ significant improvement in the fit when dropping the mean reversion term

$\Rightarrow$ **non mean reverting affine processes seem appropriate to describe the intensity of mortality**

Differences between $S_x(t)$ and ${}_tp_x$ in the non mean reverting processes



Notice the difference in the scale w.r.t. the previous graphs: this allows us to use interchangeably all the four models.

Sensitivity analysis

Evidence from the calibration of the non mean reverting processes:

- low or null diffusion parameter (σ)
- improvements of fit when adding a jump component

In the case of projected tables, maybe the low value of σ is due to the fact that the calibration is done on mortality tables that are constructed with deterministic algorithms. However, with the observed mortality tables this explanation cannot apply.

Fact: this seems to suggest that the future evolution of intensity of mortality for a head aged x presents low variability.

This does not need to be true for the future. So, what would be the effect of higher variability in the intensity λ on the survival probabilities?

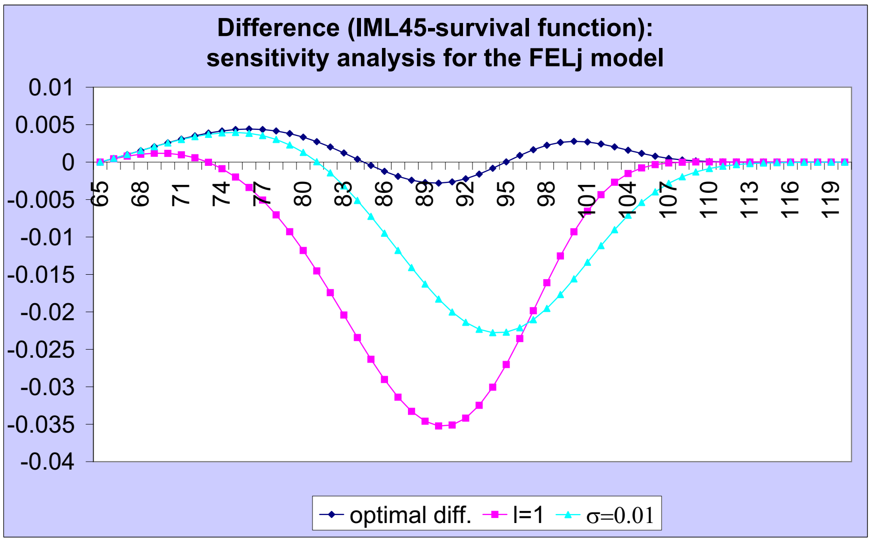
Assessing the effect of higher variability

For the processes OU, OUj and FEL, we can answer this question with analytical results: when we increase the diffusion coefficient or the jump intensity (the latter meaning a reduction in the expected arrival time of jumps), these models predict a higher survivorship.

For the model FELj analytical results cannot be obtained and we provide a sensitivity or stress test analysis. Let us consider the differences between the survival probabilities of the table and those implied by the model

- under the optimal parameter values (optimal diff.)

- with a diffusion coefficient σ and an intensity l equal to a thousand times the optimal ones.



Increasing the diffusion coefficient σ or the jump intensity l, leads to differences becoming more negative.

$$\Rightarrow$$

Increasing the stochastic part of the intensity process implies improvement in the survival probability

The link with existing models for the force of mortality

What is the relation between the stochastic intensity of mortality and the deterministic force of mortality?

$$\mu_x = \lim_{h \to 0} \frac{P(x < T_0 \leq x + h | T_0 > x)}{h}$$

In our case, we have:

$$\mu_x = \lim_{h \to 0} \frac{1}{h} \left(1 - \frac{S(x+h)}{S(x)} \right) = -\alpha'(x) - \lambda_0(0)\beta'(x)$$

For example, in the OU model the force of mortality, becomes:

$$\mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax} - 1)^2$$

If $\sigma = 0$ we have:

$$\mu_x = \lambda_0(0)e^{ax} = \lambda_0(x)$$

i.e. in this case the force of mortality coincides with the intensity of mortality for a new born individual after x years. Furthermore, the force of mortality is of the **Gompertz type**. This is straightforward also observing that if $\sigma = 0$ the evolution of $\lambda_0(t)$ is deterministic and given by

$$d\lambda_0(t) = a\lambda_0(t)dt$$

The coincidence between intensity of mortality and force of mortality is clearly no longer true when the intensity is stochastic. Furthermore:

$$\mu_x < E(\lambda_0(x))$$

In other words, the force of mortality decreases, hence the survivorship improves, when the diffusion coefficient increases.

With the other three models, we have:

$$OUj \qquad \mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax}-1)^2 - \frac{l}{a-\mu}\left(1 - \frac{a\mu e^{ax}}{a-\mu+\mu e^{ax}}\right)$$

$$FEL \qquad \mu_x = \frac{4\lambda_0(0)b^2e^{bx}}{[(a+b)+(b-a)e^{bx}]^2}$$

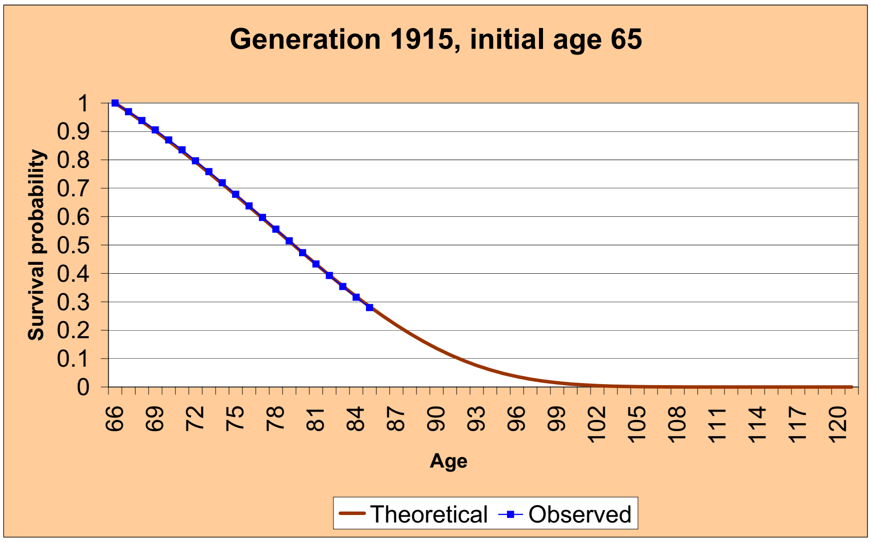
$$FELj \qquad \mu_x = \frac{4\lambda_0(0)b^2e^{bx}}{[(a+b)+(b-a)e^{bx}]^2} + \frac{l\mu(1-e^{bx})}{\mu-c-(d+\mu)e^{bx}}$$

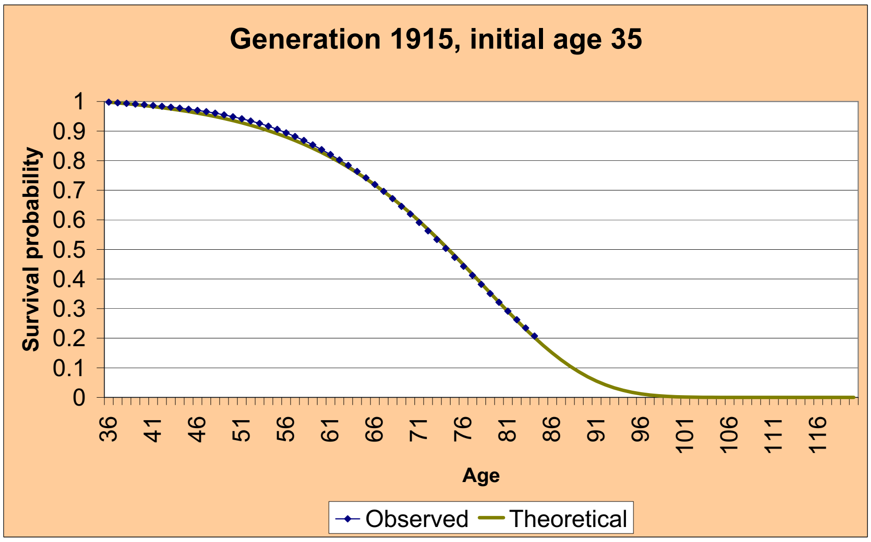
It is clear (and easy to check) that also with these three models, when the coefficients σ and l of the random part are set to 0 there is coincidence between intensity of mortality and force of mortality, which turns out to be of the Gompertz type.

Forecasting mortality

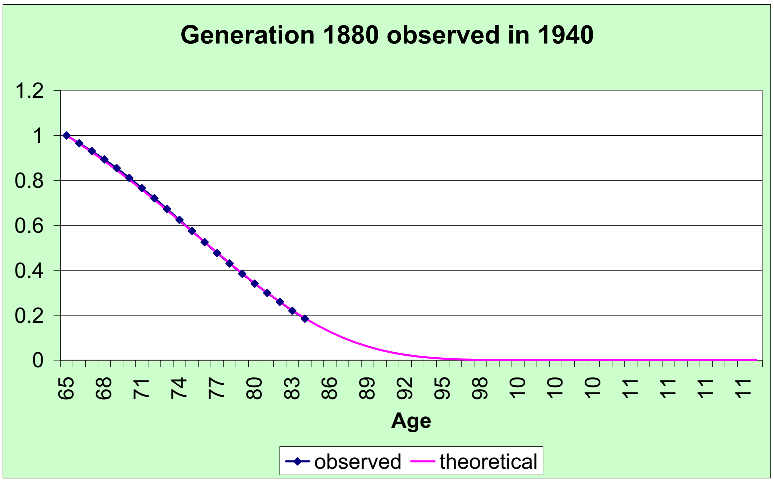
One can use this model to see what is the future evolution of mortality for a given generation.

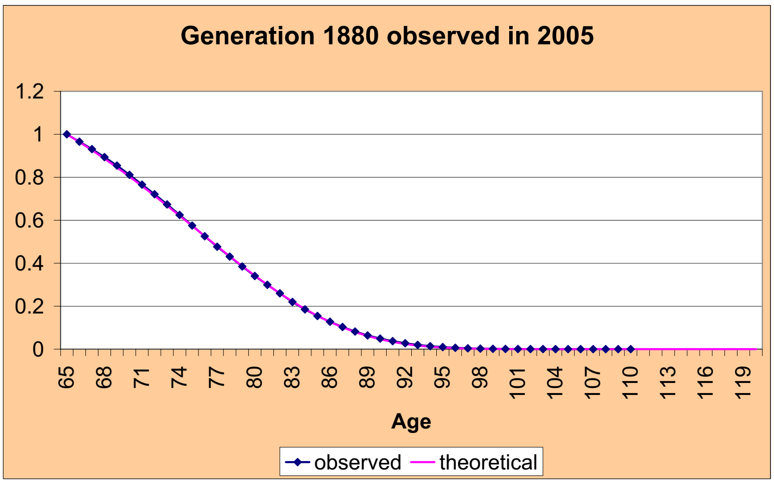
In the next two graphs, we report the mortality forecast for the generation 1915 for initial ages 35 and 65 (FELj model). The right tail of the "Theoretical" curve gives the forecast of the survival function beyond the observation date.


Generation 1915, initial age 65
Survival probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
66 69 72 75 78 81 84 87 90 93 96 99 102 105 108 111 114 117 120
Age
Theoretical
Observed


Generation 1915, initial age 35
Survival probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116
Age
Observed
Theoretical

We have applied the same forecast procedure on the generation 1880, initial age 65, in order to compare the forecasted mortality with the experienced one.





Mortality trend

Let us consider the intensity of mortality for a given initial age x and different generations. A complete description of the intensity surface would be given by a two parameters-family $\lambda_{x,gen}$ (Biffis and Millossovich, 2005).

Here we focus only on the change of generation and omit the initial age x. We have a family of intensity processes:

$$d\lambda_{gen}(t) = f_{gen}(\lambda_{gen}(t))dt + g_{gen}(\lambda_{gen}(t))dW(t) + dJ_{gen}(t)$$

where the index gen refers to the year of birth (eg 1880, 1905).

The change in $\lambda_{gen}(0)$ and in the parameters that characterize f_{gen} and g_{gen} gives the description of the mortality trend in our setting.

Mortality trend - first approach

- calibration for the sixteen generations born in years 1900 to 1915
- initial age $x = 65$
- FELj model
- for each generation we calculate the value of $\lambda_{gen}(0)$ and find a set of optimal parameters:

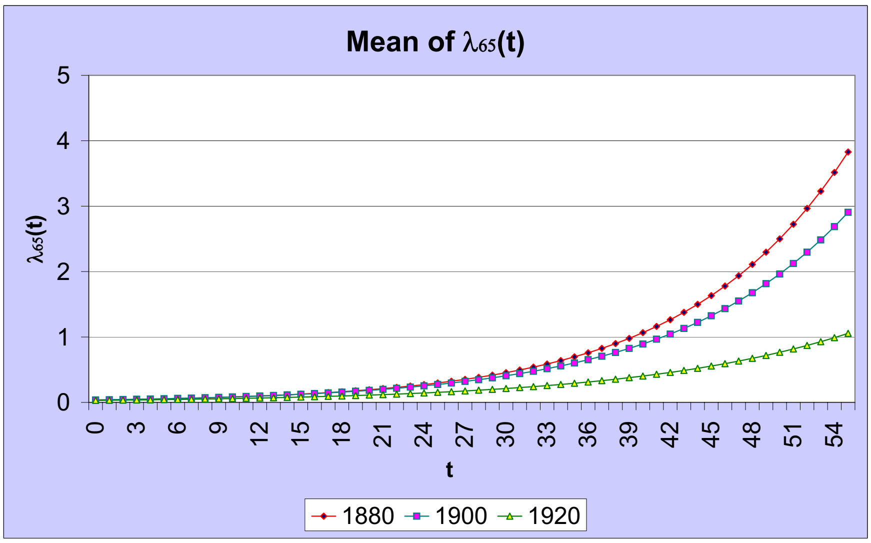
$$a_{gen}, \sigma_{gen}, l_{gen}, \mu_{gen}, error_{gen}$$

Results

- decreasing trend of a and linearly decreasing trend of $\lambda_{65}(0)$ (R^2 of 0.912 wrt to calendar year)
- the calibration errors are very small.

Mortality trend - second approach

We have simulated the process $\lambda_{65}(t)$ for the generations 1880, 1900 and 1920. The graph reports the mean of $\lambda_{65}(t)$ for the three generations. As expected, the older the generation, the higher the mean of the intensity.



Concluding remarks

- In this paper, we have described the the time of death as a doubly stochastic stopping time: namely, as a jump time whose intensity is stochastic. The intensity has been described as an affine process, with mean reversion and without it. For both specifications, the survival probabilities have been provided in closed form.

- The intensity processes have been calibrated to the UK population, using observed mortality tables for old generations and projected tables for younger ones.

- Results seem to suggest that, in spite of their popularity in the financial context, mean reverting processes are not suitable for describing the death intensity of individuals.

- On the contrary, affine processes whose deterministic part increases exponentially seem to be appropriate for describing the intensity of mortality.

- The non mean reverting affine processes proposed can be considered natural extensions of the Gompertz model.

- The stochastic component of the intensity processes seems to be appropriately described by negative jumps together with diffusions.

- Stress analysis and analytical results indicate that increasing the randomness of the intensity processes results in improvements in survivorship.

- We provide a procedure for mortality forecasting and mortality trend assessment: comparison of forecasted and experienced mortality for old generations gives very satisfactory results.