Review of the diploma thesis

Author: Bc. Kseniya Kuzminskaya

Title: Acceleration of calculations in life insurance

Opponent: Dr. Sebastiano Vitali, Ph.D.

The author analyzes the computation of the fair present value of an insurance contract. The author investigates three possible approaches: the standard "policy-by-policy" approach, the proxy function approach and the cluster approach. The last two methods are very promising for accelerating the computation of the fair value of the insurance policy. Despite the fact that the thesis should be evaluated because of its content, still the length appears to be inadequate to completely and deeply analyze the effect of such methodologies. Moreover, extremely dangerous errors have been found which can make the whole discussion of the results completely non sense. In particular, the analyzed input data are probably wrong and therefore also the estimations and the further computations are wrong.

I do not think such thesis fulfill the standards for a Master Thesis at Charles University.

Here, I list in more detail my comments and doubts:

The thesis would require a complete English correction. There are many typos and grammar mistakes. I will only list the main I found in the Introduction section, but a strict proofreading should be applied to the whole text.

insurance companies take → insurance companies consume

cash flow → cash flows

is based → discusses (focuses on)

of said processes → of these processes (or of mentioned processes)

Among the actuarial tasks belong risk management, … →this is not an English construction

requires → require

For every contract is calculated cash flow with some possible scenarios →this is not an English construction

Because → Since

the work on → the results of

We will to simulate → either "we will simulate" or "we want to simulate"

chose → choose

FV_t is the Fair Value at time t, but then the author uses t also within the formula to compute its value. I would rather use another notation, for instance: $FV_{t_0} = \sum_{t=t_0}^{T} \frac{CF_t}{(1+rfr_t)^{t-t_0}}$.. Moreover, the risk free at time t is used but it is clearly not assumed to be constant, so the present value should be computed considering all yearly risk free rate from t_0 to t.

The title of 1.3 should change from Assumptions to Definitions.

The formula at the end of page 6 would require a deeper explanation: why the fees are applied only on lapses and not on deaths and maturities? What is the coefficient q_x? I can understand is the death probability but should be clearly specified. And what is $wthd_t$? The step from second row to third row is

very unclear, and what is RP_t? The motivation for the inclusion of the interest rate i (I assume it is an interest rate because it is not specified) is not explained at all.

In section 2.2, I would use another notation because function f has already been used as aggregating function of interest rates in section 2.1.

In section 3.1, the SDE is from reference 9? Point 3 of the Wiener process definition: what does it mean $\forall 0 =$?

Why the non negative interest rates are an advantage? Nowadays, we observe also negative interest rates.

In section 4.1, the data considered from reference 13 and depicted in Figure 4.1 are not the 1Y yield curve, but the series of the beta0 coefficient of the Svensson model adopted by the ECB. The correct source of the 1Y yield is this:

http://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=165.YC.B.U2.EUR.4F.G_N_A.SV_C_YM.SR_1Y

I'm very worried that all further computations are wrong since the data used for the estimation are incorrect. Unless the author can justified why the beta0 coefficient is used as proxy of the 1Y return (I cannot see any justification). Moreover, from the 1Y data is clear that in the considered period the rate is negative, therefore, I afraid that the CIR model is not suitable.

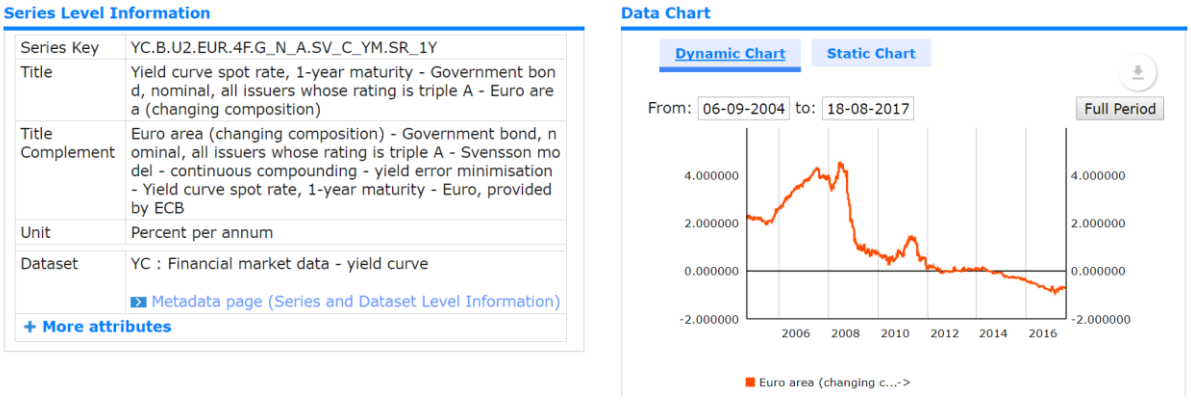

Figure 4.2: I do not understand how the histogram assigns a probability to values greater than 2.4 and lower to 1.8 since the are no scenarios going to such extreme values. Moreover, I'm also wondering if the estimations are correct, since the high volatility of the historical series (even if wrong) which spans from a maximum value of 2.1 to a minimum value of 0.7 is not fully represented by the generated scenarios which spans from 1.9 to 2.4. Another issue is the initial point: it is not clear to me which initial r_0 has been considered for the scenario generation: it should be the last observation of the historical series (approx. 1.5) but it seems different (approx. 2.2).

Results: to me it is not clear the meaning of "model points". Sometimes it seems to be each realization of the stochastic rate generator (cf. end of Introduction), sometimes the different policies we are considering (cf. beginning of Section 2.2). In Tables 4.1 and 4.2, the mean of errors could be misleading because there could be compensation between positive and negative errors: it would be better to include also some volatility of errors or max error to give a better idea of the quality of the approximation. Moreover, it is not clear under which assumption of interest rates the author calculates the policy-by-policy value in order to compare with the analytic function value. The doubt comes from the fact that a 0.000% error is very strange considering that only 20 scenarios have been adopted for the discretization of the interest rate diffusion process. Therefore, I would expect a larger difference between the two and a sensitivity analysis according to the cardinality of the scenario set. Such results look suspicious.

In Prague, 4.9.2017 Dr. Sebastiano Vitali, Ph.D.