Some remarks about forecasting longevity

Emilio Venezian

First of all

- Thank you!
- I am honored to have been invited.
- In some ways I am an outsider to life insurance, it is not my specialty.
- Outsiders are not often useful.
- Sometimes, however, they have new ideas that they bring from other fields.
- I hope that what I have to contribute will be worth the time you invest on listening to it.

What I hope to do is

1. Try to focus on whether we are trying to forecast period longevity or cohort longevity.
2. Raise some questions about what is being done.
3. Provide you with some of my feelings about the quality of the data typically have as a starting point.
4. Propose a way to measure of how good a forecast is.
5. Report on what happened when I engaged in performing such a test.
6. Raise some questions about what I have done.

- And I have a limited time budget.

1. Why are we trying to forecast longevity or mortality?

- Our interest is related to some real economic activity.
- If you were in the business of arranging funerals or making supplies for them I might talk about period mortalities.
- Insurers might also be interested in period mortality from the perspective of planning cash flow and liquidity.
- Much more important, however, are the interests of insurers and their regulators in making sure that contracts are priced correctly and reserves are set adequately.
- In insurance we may be trying to price a life insurance contract (emphasize mortality) or an annuity contract (emphasize longevity).

We need to price products and set reserves for future losses **now**

- We do not know the year in which the applicant, annuitant, or insured is going to die.
- We do have information on the year of birth.
- Our focus should be on forecasting cohort mortalities.
- The vast majority of published papers on forecasting does not even state clearly what mortality is being discussed; those are probably using period mortalities.
- And they forecast period mortalities for the next 50 years or more from data that we have available now.
- That may mean period mortality data that is no more recent that five or ten years ago, depending largely on the size of the population.

We have little data on those cohorts.

- To forecast mortality for those cohorts from historical cohort mortality data we have to extrapolate very far; that usually results in large forecast errors.
- Using period mortality we can have access to recent data and do not need to expose ourselves to the large forecast errors.
- But we do expose ourselves to bias.

Period and cohort mortality coincide only in a static world.

- The world in which we live is not static: mortality changes over time.
- In fact, in most countries mortality has been declining over the ages.
- The most recent period mortality table will tend to overestimate life premiums and underestimate annuity premiums.
- So we end up with profits we can hide and deficits that we cannot hide.
- We have big problems with longevity.

We have an empirical issue

- Are the errors involved in forecasting from the cohort data over long periods of time worse than those involved in using more recent period mortality and having a bias?
- That, of course, is a question that should be answerable.
- It turns out to be more difficult than I anticipated.

2. What I hear and read

- Suggests that we have two groups of people.
- One group works with period data.
- The other group works with cohort data.
- There seems to be little interaction between these groups.
- This lack of interaction appears evident when you look for data such as life tables.
- The protocols discuss a lot of detail about how the period life tables are calculated and how the cohort life tables are calculated.
- But nothing much is said about whether the two sets are **reasonably** consistent.

Perfect consistency is impossible

- Unless you have infinitely large homogeneous populations and infinitesimally small time segments.
- The problem is we do not have even the semblance of consistency.
- I chose to use data from the Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de
- HMD smooths the period life tables but does not smooth the cohort life tables.
- That makes life difficult when you want to ask the question I propose to address.

3. I will use life tables, but they are too little and too late

- If forecasting is to be a meaningful, as opposed to a purely academic endeavor, we should go beyond the data usually provided in life tables.
- Life tables involve assumptions, often arbitrary, about so many things that it would take all my time just to go through them.
  - Some of them we can do little to fix, but we could state them:
    - the completeness of death registration or reporting
    - the extent and impact of "medical tourism"
    - the timing of entry and exit of people through migration
  - Others are just silly adherence to published work or received dogma and could easily be improved, among them:
    - the timing of births within a year,
    - the timing of deaths within a year, and
    - the age distribution within a cohort.
- We will not know whether these differences will be material, my guess is they will matter a little at age 0 and a lot at high mortalities.

Moreover

- The life tables conceal a great deal of importance if we want to perform sound statistical analyses.
- In particular, that do not reveal how reliable the underlying the numerators and denominators are so the ratios can be weighted appropriately.
- In my view
  - It is just silly to give the same weight to a ratio based on a numerator of one and a denominator of two as to one based on a numerator of 100,000 and a denominator if 200,000.
  - It is equally silly to report no deaths in an age group with five person years as a probability of death (or a mortality) of zero.
- Life tables would be much more useful if they had measures of the reliability attached to the numbers.

What do you mean by "death rate" or "mortality rate"?

- In an infinite, closed population with $P(a, t_0)$ people alive at age a at time t_0, the number of people alive at time t_1 is usually expressed in one of two ways

1. $P(a + t_1 - t_0) = \big(1 - q(t_0, t_1)\big)P(a, t_0)$, or
2. $P(a + t_1 - t_0) = e^{-\mu(a, t_0, t_1)}P(a, t_0)$

- but other conventions also exist.
- Tables are not always clear as to what is mean by the terms used.
- That can create big problems.

I will use data from HMD

- When I started that sounded like a good enough source for my purposes.
- I downloaded the cMx and Mx tables on May 2, 2011.
- These gave "death rates" for cohort and periods, respectively.
- I saw several problems, so on May 11, 2011 I wrote to the director of the HMD project outlining the major problem: the meaning of "deaths rates" seemed to vary from country to country and between cohort tables and period tables.
- Some countries had death rates that exceeded one in both period and cohort tables, some had them in the cohort table but not the period table, the U.S. had no value greater than one in either table.
- Going through the history would take more time than I am allowed.
- On June 4th the HMD websites were unavailable, on June 5th I reached them and found many changes.

Any analysis is complicated…but

- If we let X denote either cohort or period mortality, a denote the age, and t_X denote the cohort or time of death then the approximate relations

$$ {}_Xq(a,t_X) = 1 - e^{-{}_X\mu(a,t_X)} $$
$$ {}_X\mu(a,t_X) = -ln(1 - {}_Xq(a,t_X)) $$

  should bring these into reasonable agreement.
- The adjustment will not be perfect because people aged a dying in year y may belong to different cohorts.
- The differences should be smaller than those observed not adjusting to the same convention.
- moreover, we should have

$$ {}_Pq(a,a+c) \cong {}_Cq(a,c), \text{ and} $$

$$ {}_P\mu(a,a+c) \cong {}_C\mu(a,c) $$

- These relations would hold exactly if we had infinite populations and data at infinitesimal intervals…but then life would not be long enough to look at the data.

I tested these relations on the new data

- The relations between q and μ did not work well at all.
- The relations between ${}_Pq(a,a+c)$ and ${}_Cq(a,c)$ worked better than those between the corresponding values of μ in most, but not all countries.
- So I based my analyses on the q values of the cohort and period life tables.
- So even in highly managed data bases we need to be very careful with our data and examine it carefully before we act.
- And even more carefully before we draw conclusions.
- One other important change: the May data provided numbers for cohorts that were not complete, the new cohort life tables included only "almost complete" cohorts.

The problems are probably due to "smoothing"

- The practice of smoothing creates many problems even if it is done with great care.
- The foremost is that the user does not know to what extent the smoothing reflects the biases or preferences of the smother.
- But almost as important, there is no way of knowing whether consistency, exact or approximate has been maintained.
- In particular, HMD forces all **mortality** (not just probability of death) values in the period tables to be less than or equal to one.
- No such restriction is imposed on the cohort values.
- So we have some problems right from the outset.

3. Designing a fair test

- To be fair the test must avoid using, as far as is possible, any information about the structure of mortality rates.
- Forecasting 100 years into the future you run the risk of negative forecasts.
- I forecast the logarithm of μ . That avoids bounding the forecast in [0,1] (as for q) or [0,∞] (as for μ).
- I assume that at each age the logarithm of mortality for cohorts (periods) differs from that of previous cohorts (periods) because of random changes in level and slope.
- Wars and epidemics typically affect the level but not the slope, sanitation and health care technology can affect both.
- Given the assumptions, exponential smoothing is a suitable tool.

The next issue is choosing smoothing constants.

- I do not want to choose those optimally.
- A reasonable rule in the absence of optimization over a "planning period" is to select constants of 0.9 (an effective averaging period of about 10 years) for both level and slope.
- I can then apply the method to both cohort mortalities and period mortalities.
- I assume that at any point the data available is that relating to events that took place two years before the current year.
- The HMD database does report mortalities of zero.
- Since that makes the log equal to negative infinity I modified the usual equations so when that value was found it was replaced by the forecast value for that period calculated on the previous year.

That leaves open the issue of a figure of merit

- The usual figures of merit come from statistics.
- They include the average bias, the standard deviation about the "true" value, the standard deviation about the mean.
- Are we going to apply those to the logarithm of mortality, the mortality itself, or the probability of survival?
- Or can we find a better figure of merit?

If we emphasize longevity

- If our main concern is with funding annuities we could base our figure of merit on how well forecast premiums compare with "actual" premiums.
- Since the focus is on longevity; I think an appropriate contract would be a single premium deferred annuity (SPDA) purchased at age *y* at the time the forecast is made and pays $1 in benefits at every anniversary starting at some age such as 60 and continuing until age 95 provided the annuitant is alive.

Why 95?

- We need to stop somewhere because we need a reasonable number of "complete" cohort tables extending from birth to the maximum age to be able to compute the "actual prices".
- There are not many countries in the HMD that can satisfy that requirement with a maximum age of 95.
- In the May data I had found only five that I was willing to use.
- In the June data there are only two because partial cohorts are not reported.

But I have not yet specified a figure of merit in full

- So let us think about the price for a contract such as I have specified.
- The future value at age 60 of the payment made at age y has to be equal to the present value at age 60 of the future annuity payments.
- So we have

$$\Pi(y,c,r,m\,) = \frac{\sum_{k=60}^{95} \frac{S(y,c,m)}{(1+r)^{k-60}}}{S(y,c,m)(1+r)^{60-y}}$$

where $\Pi(y,c,r,m)$ is the price of the SPDA for cohort c at age y and interest rate r,

$S(y,c,m)$ is the fraction of the population in cohort c surviving to age y.

- These can be computed to for the actual cohorts (m=1), the forecasts based on cohort data (m=2), and the forecasts based of period data (m=3).

We could make that into a figure of merit

- For example, we could consider averaging ratio $\frac{\Pi(y,c,r,1\,)-\Pi(y,c,r,i\,)}{\Pi(y,c,r,1\,)}$ over all cohorts available.
- I prefer the ratio of the difference $\Pi(y,c,r,1\,)-\Pi(y,c,r,i\,)$ to the derivative $\Pi'(y,c,r,1\,)$ expressed in term of dollars of premium per basis point in the assumed discount rate.
- It reflects the fact that mortality rates **are not the only forecast called for in pricing**.
- It provides information about whether more effort should be allocated to forecasting longevity, interest rates, or incorporating stochastic models of interest rates in the pricing equation.

That "statistic"

- Gives a first order approximation to the number of basis points away from the "expected" interest rate that need to be applied in order to make the forecast rate "correct"
- Assume the expected rate of return is 5 percent, a number I will use for purposes of illustration.
  - Then a value of +19.3 means the policy would "break even" if it were offered with an interest rate of 5.193.
  - And if the value were -20.7 the policy would not break even if its is offered at any interest rate above 4.793.
- So that seems a decent way of measuring accuracy.
- It gives us information as to where more effort is needed.
- Both the average and the standard deviation of the measure contain valuable information.

One more caveat

- The HMD cohort life tables contain fictitious data because they include estimates for "almost extinct cohorts".
- That would bias the comparison in favor of forecasting from cohort data.
- I therefore eliminated all data on cells for which the date of death was later than that of the most recent period mortality table for the country.
- On June 14th I received a reply from HMD that said "thanks to you, we found a couple of bugs in our program and we have been working on fixing them."
- I was promised an answer to other questions (such as why there were mortalities of zero in the "smoothed" period data) in a month or so.
- The answer came at the end of June: "because there were no deaths."

5. Actually doing it

- Since the format of HMD had changed I now used the gender specific period and cohort life tables.
- In the June data there were only two countries (France and Sweden) had enough data, others did not have any years with cohort data from ages 0 to 95.
- I think the results are interesting enough to discuss.
- They are not unlike the results obtained from the more questionable May data.
- I always assumed that the most recent data available was two years before the calendar year at issue.

Two important cautions

1. As in all cases of forecasting, what has worked in the past many not work in the future.
2. I am reporting on how the well pure rates using forecasted future mortality compared with the future rates in given in the database, not how they may compare to reality.

- However, we have no other way of estimating how good or bad forecasts are likely to be.

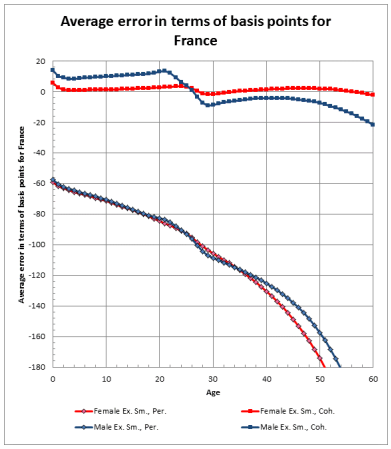
Analysis of bias in terms of basis points for rates for all ages from 0 to 95							
	Number of			Bias in basis points			
	Years	Points		Female		Male	
France	8	768		Period	Cohort	Period	Cohort
			Ave.	-155.15	0.95	-156.89	-2.65
			SD(Ave.)	13.27	0.38	13.87	0.51
			p(bias = 0)	3.39E-29	1.20E-02	1.58E-27	3.17E-07
Sweden	72	6912	Ave.	-219.51	-11.18	-217.62	-4.30
			SD(Ave.)	3.46	0.70	3.46	0.82
			p(bias = 0)	0.00E+00	2.06E-56	0.00E+00	1.86E-07

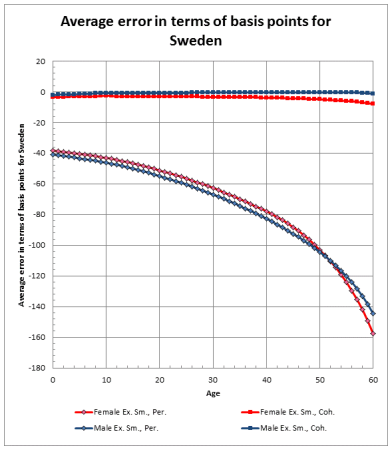
Analysis of bias in terms of basis points for rates for all ages from 0 to 60							
	Number of			Bias in basis points			
	Years	Points		Female		Male	
France	8	488		Period	Cohort	Period	Cohort
			Ave.	-45.35	0.43	-69.07	-0.78
			SD(Ave.)	22.58	0.60	6.74	1.30
			p(bias = 0)	4.49E-02	4.67E-01	3.68E-23	5.49E-01
Sweden	72	4392	Ave.	-67.01	-3.67	-69.07	-0.78
			SD(Ave.)	6.78	1.12	6.74	1.30
			p(bias = 0)	6.61E-23	1.07E-03	1.92E-24	5.49E-01

Some results

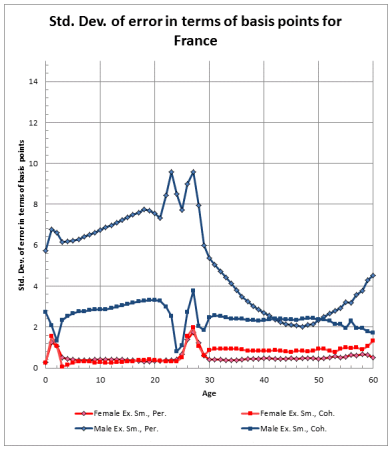
- I will present some results in graphical form in the next several slides.
- These are averages and standard deviations of the biases.
- The graphs re-emphasize the more global results presented in the tables.

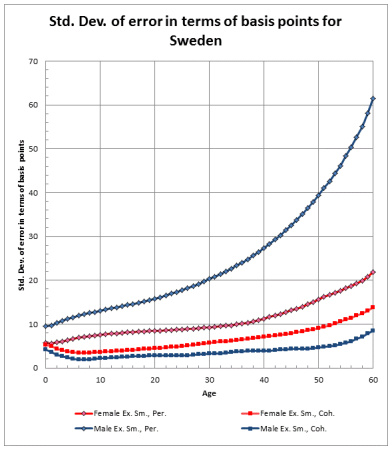
Average error as a function of age



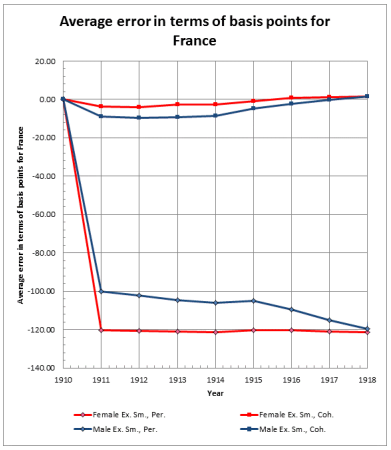


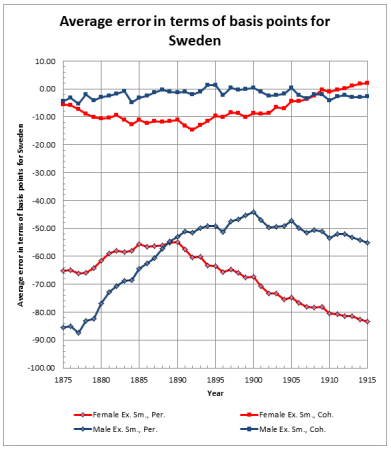
Standard deviation of error as a function of time





Average error as a function of time



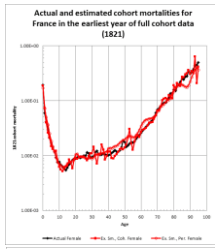


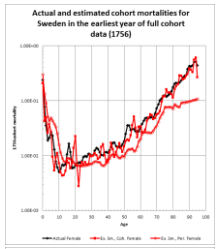
Those results seem just too good

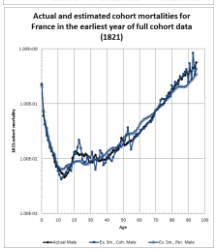
- Of course a sample of two countries is not very convincing.
- But I keep wondering whether it is just the way the HMD tables are calculated or whether it is real.
- If it is real, then using information other than past rates (such as prevalence of smoking or overweight) should lead to even better predictions.

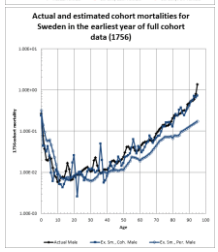
Averages are averages

- But sometimes individual sets are interesting.
- The following slides show three sets of curves for each country
  - cohort mortalities for the first year with complete data for ages 0-95 in the cohort mortality,
  - cohort mortalities for the last year of any cohort data, and
  - period mortalities for the most recent year with period mortality data
- I will show panels for females and males. Each panel has three curves:
  - the actual values (black lines)
  - forecast values based on cohort data (color lines with solid square points), and
  - forecast values based on period data (color lines with empty round points).


Actual and estimated cohort mortalities for France in the earliest year of full cohort data (1821)

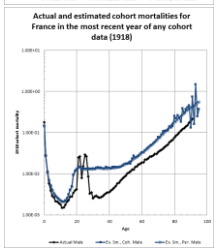

Actual and estimated cohort mortalities for Sweden in the earliest year of full cohort data (1756)


Actual and estimated cohort mortalities for France in the earliest year of full cohort data (1821)

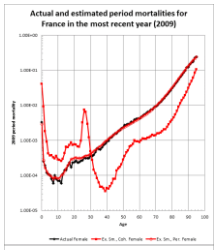

Actual and estimated cohort mortalities for Sweden in the earliest year of full cohort data (1756)

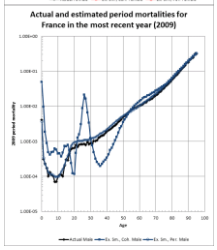

Actual and estimated cohort mortalities for Sweden in the most recent year of any cohort data (1917)


Actual and estimated cohort mortalities for Sweden in the most recent year of any cohort data (1917)


Actual and estimated cohort mortalities for France in the most recent year of any cohort data (1918)


Actual and estimated cohort mortalities for Sweden in the most recent year of any cohort data (1917)





Tentative conclusions

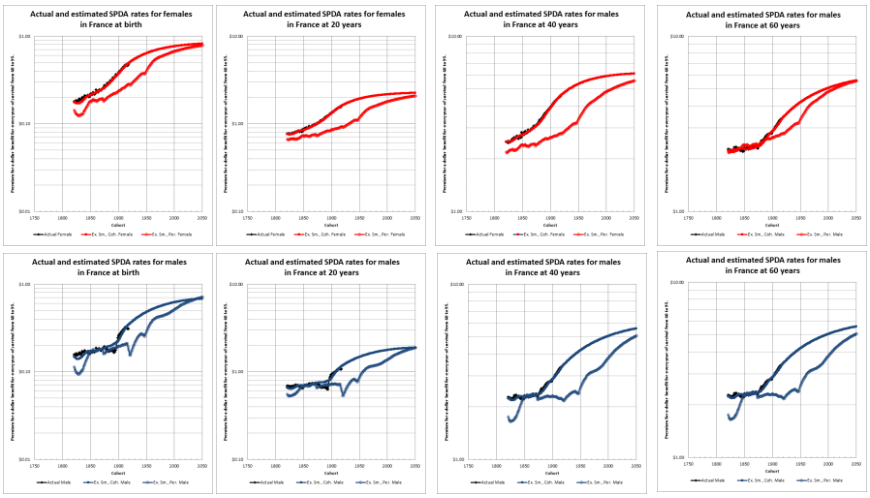
- Please recall that, in spite of everything I could do, the data still presents problems.
- So does the analysis: the exponential smoothing was performed on the logarithm of mortality and the mortality was then inferred by taking the exponential of the smoothed logarithm.
- Jensen's inequality tells us that the result is biased; the mortalities should be somewhat higher than what this calculation yields.
- That may or may not account for some of the results if we are viewing this as a purely academic exercise.
- As a practical matter this can be considered just a property of the methodology.

Cohort forecasting seems better

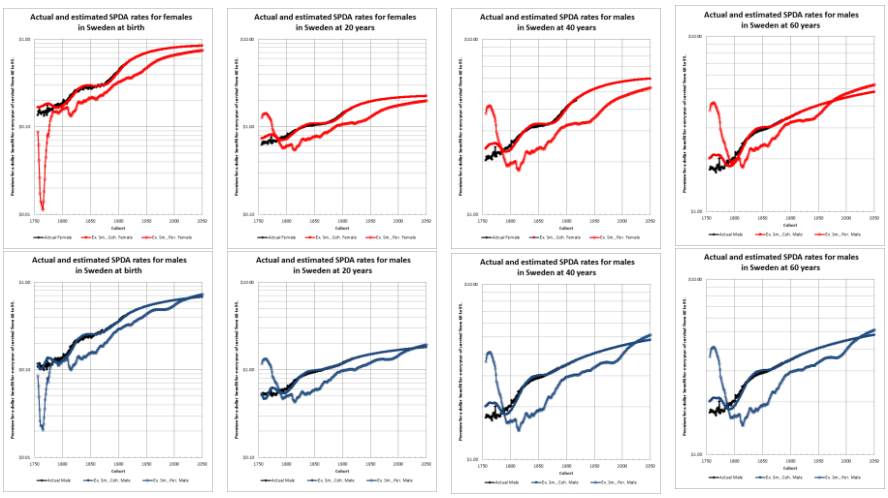
- One matter may be settled, though I do not have the time to show you the data: exponential smoothing of cohort data is considerably better than using the period mortality of two years earlier (with no forecasting) as an estimator of cohort mortality.
- However, all this applies to longevity. Things might work differently if we are interested in pricing life coverages, and the signs of the biases would most probably be reversed.
- That is not good if you are selling life insurance.

Just for fun: what are the forecasts for rates well into the future?

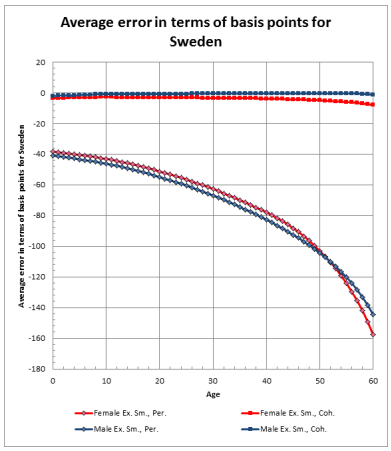
France

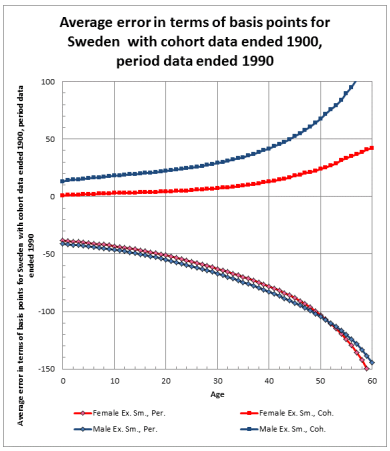


Sweden

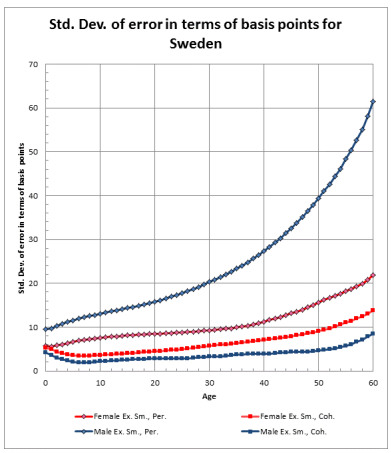


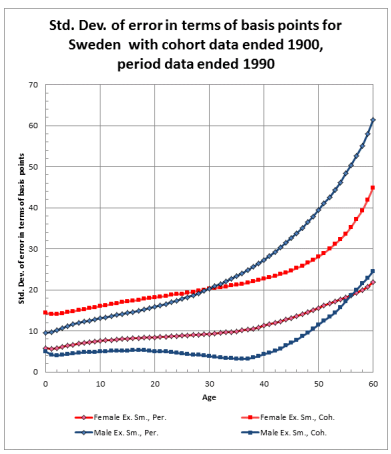
Here is what happens if you throw away recent data



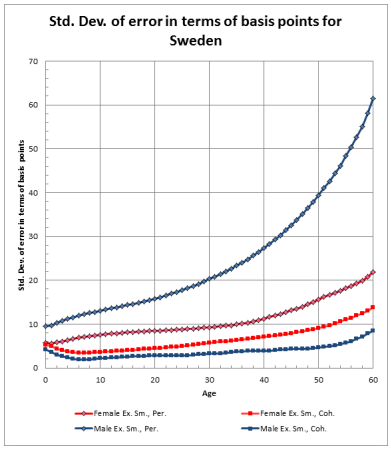


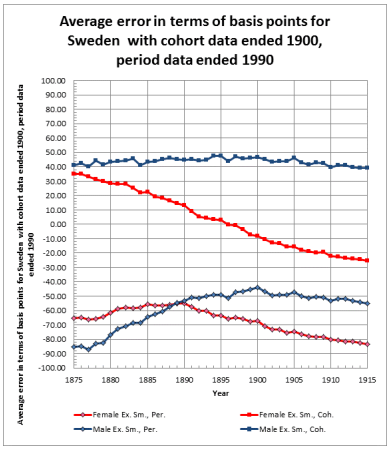
and the standard deviation by age





and the average error by year





Refinement may not help

- The biases found by naïve exponential smoothing are of the order of 10 basis points.
- If we are going to put a lot of effort into reducing that error we should consider the errors involved in assuming that the interest rate forecasts are not good
- And discounting as though the interest rate is certain and constant over the life of the contract is also questionable.
- Should we really bet good money on the proposition that we know today what the interest rate will be for the next 30 to 50 years to within one tenth of one percentage point?

6. There is still a lot of work to do

- The present value calculation involves the interest rate in the denominator.
- So Jensen's inequality gets in the way again.
- We may need a good stochastic model of interest rates as a function of current rates so we estimate the price that reflects the expected value of the revenues that would accrue from investment of the balances and the costs that arise because of longevity.
- Or perhaps the incorporation of future and forward rates into the calculation.
- My guess is that those effects are at least as large as those of misestimated longevity.
- That makes the development of usable interest rate models a priority.

Also

- I have confined myself strictly to forecasting as "casting the data ahead".
- One might want to use information about
  - individual variables such as smoking habits or obesity,
  - social trends in such variables, or
  - cohort and year effects.
- And while this may help with the problems of pricing and reserving it may not help with the common social problem:
  - all I have done relates to "stable" populations,
  - real populations have variable birth rates and jagged age profiles,
  - the problem of "pay-as-you-go" pension systems remains a problem since it requires forecasts of the future population of contributors.
- Dealing with that problem requires long-term forecasts of the rate of household formation and the birth rate.

Two final thoughts

One

- We are devoting a great deal of effort to longevity, perhaps because we believe we have lots of data.
  - The social problem goes beyond that, since it often includes providing for the costs of disability.
  - Those costs are not small, and we should not neglect them.
  - Even the private market has to deal with disability insurance.
  - Maybe we should change the language to remind us of a part of the field that needs cultivation and data.

Two

- I have said nothing about longevity risk.
- That is a very popular topic in academic journals.
- People seem to forget that the risk can be hedged by balancing annuity risks with mortality risks.
- I suspect that is more efficient than securitization but I have not seen papers on the subject.
- And securitizing all your mortality risk is not efficient if you also have longevity risk (and *vice versa*).
- So I invite you to think about hedge ratios when you have portfolios of life and annuity contracts with different features.

Thank you

- Your comments are welcome at

  emilio_venezian.yahoo.com

- I hope to have a more complete copy of this work posted at

  http://drvenezian.wordpress.com

  in the next month or so.