Mortality Forecasting at Advanced Ages - Applying the Lee-Carter Model to an Economic Panel

Malene Kallestrup-Lamb

CREATES - School of Economics and Management, Aarhus University

Visiting Post-Doctoral Fellow at Cass Business School

Seventh International Longevity Risk and Capital Markets Solutions Conference, Frankfurt.

September 2011

Motivation

- Over the last century, we have seen a steady decline in mortality rates - ultimately leading to longer life expectancies.
- These improvements have not been constant and there have been differences both between ages, genders and countries.
- In general, increased life expectancies are desirable to the individual.
- Problems arise for governments and pension funds due to the uncertainty about how long an insured individual will actually live.
  - They have both issued contracts that guarantee individuals a lifelong income stream independently of how long they live.
- This increase the pressure on pension funds and government expenses through commitments to early and old age pension, healthcare, and caretaking in retirement homes.

Motivation

- Further research is needed in terms of identifying which underlying factors are driving differences in longevity trends such that governments and pension funds get the necessary insight to manage their longevity risk as well as quantify future costs.
- Currently, most research is based on publicly available mortality data for total populations, broken down by country and gender.
- However, as Coughlan et al. (2007) point out, population characteristics relevant to mortality also include marital status, social class, health, employment, postcode, etc.
- Moreover, as taking gender of an insured individual into account as a risk factor now is conisdered discrimination due to the recent ruling in European Court of Justice (2011), the pension industry will have to come up with new ways in addressing socio-economic differentials in longevity.

Motivation

- This paper complements the existing literature by developing more sophisticated data that gives further insight into which subpopulations are driving the improvements in life expectancies.
- Besides age and gender we also account for marital status and regional location in the mortality forecast.
- Studies have shown that single, divorced and widowed individuals have higher mortality than married individuals, see Hu and Goldman (1990), Johnson et. al. (2000), and Murphy (2007).
- The lower mortality for married individuals is explained by two main reasons.
    - The effects of selection of low-risk individuals into the marriage state.
    - The protective effects of marriage.

Motivation

- Heterogeneity in mortality rates can also arise across regions due to differences in wealth, hospital access, crime, pollution, etc.
- Richards (2008) showed how the pension industry have started improving annuity pricing by using address data.
- While pension funds have access to characteristics such as marital status, address, income, etc. within their client base, such details are generally not available for a total population.
- We exploit the fact that we have access to such data and show that differences in future marital and region specific life expectancies exist.

Literature

- The Lee and Carter (1992) method represents one of the most influential proposals regarding mortality predictions and has become the standard approach.
- They suggest a model which allows for random fluctuations in future mortality, described as a function of a single time index.
- The simplistic nature of the model has motivated a rich number of applications across industrialized countries.
  - Carter (1996), Wilmoth (1996,1998), Carter et.al. (2001), Lundström and Qvist (2004).

Literature

- Considerable work has been done to refine and extend the method.
- Lee and Miller (2001) argued for the following improvements to the original model.
  - The fitting period should be reduced to start in 1950 in order to avoid structural breaks in mortality.
  - The forecast should start with observed death rates rather than with fitted ones thus eliminating a jump between the observed and forecasted death rates in the first year of the forecast.
  - The adjustment of k should involve fitting the period life expectancy e_0 in year t.
- Booth, Maindonald, and Smith (2002)
  - Introduced procedures for selection of an optimal fitting period.
  - Proposed the adjustment of k should be done by refitting to the observed age distribution of deaths instead of the total number of deaths.

Model

- Lee and Carter (1992) suggested a log-bilinear model for mortality rates

$$\ln m_x(t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}.$$

- α_x represent age-specific constants describing the general pattern of mortality averaged over time.
- k_t known as the mortality index summarizes the change in the level of mortality over time, thus capturing the main time trend in death rates.
- β_x represents how rapidly mortality at each age varies, when the mortality index changes.
- To ensure a unique set of parameter estimates further restrictions need to be imposed. Lee and Carter propose the following normalization (other normalizations are also possible):

$$\sum_{t=t_1}^{t_n} k_t = 0, \quad \text{and} \quad \sum_{x=x_1}^{x_m} \beta_x = 1$$

Model

- As this is a regression framework with no observable quantities on the right hand side, ordinary regression methods cannot be used to estimate the model. However, an exact least square solution can be found by using the first element of the singular value decomposition (SVD), see Nash (1990).
- The parameter estimates are given as follows:

$$
\begin{aligned}
\widehat{\alpha}_x &= \frac{1}{(t_n - t_1 + 1)} \sum_{t=t_1}^{t_n} \ln m_x(t), \\
\widehat{\beta} &= \frac{\mathbf{v}_1}{\sum_{j=1}^{x_m - x_1 + 1} v_{ij}}, \quad \widehat{\mathbf{k}} = \sqrt{\lambda_1} \left(\sum_{j=1}^{x_m - x_1 + 1} v_{1j} \right) u_1
\end{aligned}
$$

- Once the k_t is estimated, standard Box-Jenkins methods are used to determine the optimal forecasting model.

Model

- In the majority of applications of the LC framework the random walk with drift, $ARIMA(0,1,0)$, is chosen: $k_t = k_{t-1} + c + \epsilon_t$ which give us following forecast: $\dot{k}_{t_n+s} = k_{t_n} + ck$.
- I have tested different specifications for k_t and significant forecast differences are found when the unit root hypothesis is rejected.
- Below we only report results using a random walk with drift, a Booth el. al. (2002) adjustment for k_t and jump-off corrected mortality rates. This implies forecasted mortality rates given by

$$\dot{\mu}_x(t_n + s) = \exp\left(\widehat{m}_x(t_n) + \widehat{\beta}_x\left(\dot{k}_{t_n+s} - \hat{k}_{t_n}\right)\right)$$

- The period life expectancies at age x in calendar year t are:

$$\begin{aligned}\overline{e}_x^{\uparrow}(t) &= \int_{\xi \geqslant 0} \exp\left(-\int_0^{\xi} \mu_{x+\eta}(t)\, d\eta\right) d\xi = \frac{1 - \exp\left(-\mu_x(t)\right)}{\mu_x(t)} \\ &\quad + \sum_{k \geqslant 1}\left[\left(\prod_{j=0}^{k-1} \exp\left(-\mu_{x+j}(t)\right)\right) \frac{1 - \exp\left(-\mu_{x+k}(t)\right)}{\mu_{x+k}(t)}\right]\end{aligned}$$

Data

- We use yearly register data directly from Statistics Denmark for the population aged 60-100 for the time period 1980-2005.
- We follow the notation in Lundström and Qvist (2004) and calculate the central death rates as

$$m_x(t) = \frac{d_x^t}{\left(P_{x-1}^{t-1} + P_x^t\right)/2}$$

where d_x^t is the number of end of year t deaths for age x and P_x^t defines the population aged x at time t.
- For every single person in Denmark, we have information about:
  - Age, gender, marital status, geographical location, labour market status, level of education, type of occupation and sector, own income, family income, wealth, pension savings,and the individual's medical history when alive, as well as time and cause of death.

Data

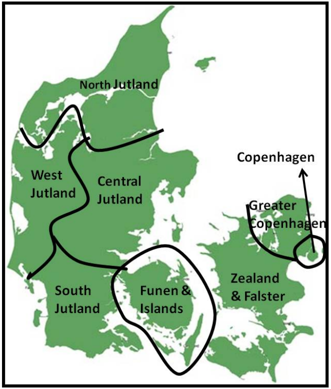
- We limit ourselves to ages below 100 years, as mortality events above that age still are relatively rare.
- We smooth observed death rates at older ages to improve the representation of the underlying mortality process by using a spline, based on Wood (1994).
- We choose 80 to be the starting age for the smoothing procedure, thus indicating that the death rates cannot decline above this age.

Data - Marital Status

- We distinguish between married and singles, where the former includes both individuals that are married or cohabitating.
  - We assign the status according to which of the two is observed the previous year and not at the time of death.
  - This should eliminate any potential concerns about the phenomenon of spouses dying close together in time thus influences the mortality rates for singles.
- We are not able to create a third category of widow and widowers, as the individuals are censored from the left.
  - We observed all individuals in the age category 60-100 starting 1981 and each year we recalculate the mortality rates for each age.
  - However, due to data restrictions we have no prior information of the individuals.
  - For instance, an individual might just have been widowed in 1979 which we do not observe or an individual who are 95 years old in 1981 might be single now but have a long martial path unobservable to us.
  - Thus creating a third category based on the current data would create biased estimates.

Data - Region

- We distinguish between eight different regions in Denmark: Copenhagen, Greater Copenhagen, Zealand & Falster, Funen & Islands, South Jutland, West Jutland, Central Jutland, and North Jutland.



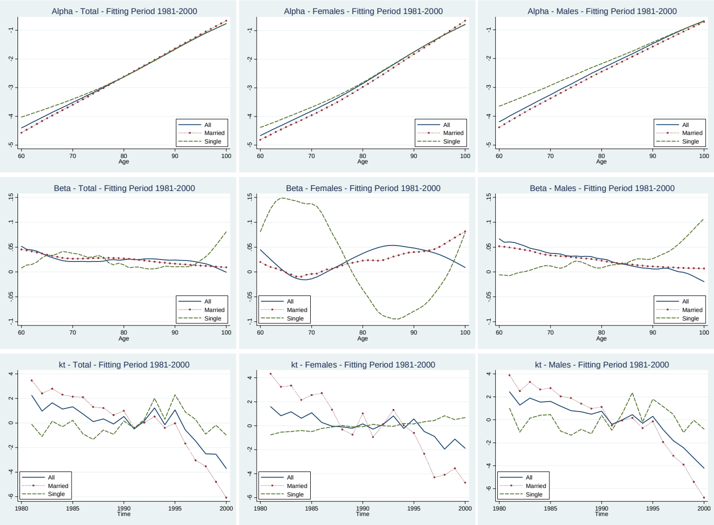
Data - Life Expectancies

- From 1900 to 1950 the growth rates in life expectancy were rather high in Denmark, amounting to an increase of around 18 years.
- During the second half of the century there was a slowdown with only a 7 year gain.
- The reason being that mortality trends have been influenced by events such as the Spanish flue, World War II, diets, The Welfare State, smoking and alcohol behavior, changes in cause of death.
- Slowdown for both males and females.

	Life expectancy at birth			Life expectancy at age 65		
	Total	Females	Males	Total	Females	Males
1900	51.95	53.54	50.22	11.58	12.08	10.95
1910	58.09	59.63	56.41	13.04	13.56	12.38
1920	57.55	58.31	56.75	12.83	13.00	12.63
1930	62.31	63.24	61.32	13.08	13.28	12.84
1940	66.27	67.21	65.30	12.67	12.93	12.38
1950	70.33	71.52	69.10	13.83	14.07	13.55
1960	72.21	73.99	70.44	14.40	15.10	13.63
1970	73.38	75.93	70.88	15.23	16.58	13.77
1980	74.16	77.18	71.17	15.72	17.55	13.63
1990	74.92	77.73	72.02	16.08	17.82	13.99
2000	76.90	79.12	74.44	16.92	18.22	15.16

- This become important as the register data is restricted in the time dimension.

Marital status



Estimates of alpha, beta & kt for the time period 1981-2000 - Total, Females & Males - All, Married & Single.

Marital status

- The α_x's exhibit the same upward trend in mortality along the age dimension.
  - However, married individuals have lower mortality rates and this difference is more pronounced for males.
- Separate estimation by gender reveals irregularities in the β's.
  - Single females in the age group 75+, with negative betas, have had a rise in mortality when it was falling for the age group 60-75. Whereas married females have experienced a slight upward trend in the age specific component indicating larger gains for the oldest.
  - For married males we find a negative sloping age component but the opposite holds for single males. This upward sloping trend indicates decreasing mortality for older ages.
- The estimated k_t's for married and singles exhibit opposite trends.
  - We see a constant or slightly positively sloping mortality index for singles, whereas for married individuals the slope is negative. This implies that the drop in mortality was much faster for married than singles.

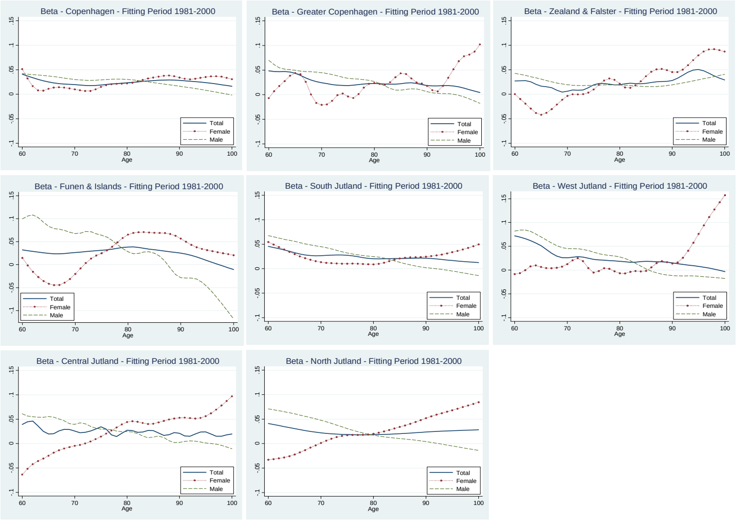
Marital status

- Females: a 60 year old married woman has a life expectancy of 24 years, whereas for a single woman it is 21.5 years in 2001. Over a 50 year time period, it increases with almost 4 years for married women whereas for singles it is only one year.
- Males: a 60 year old married man has a life expectancy of 21 years but only 16 years for a single man. In 2050, we see an increase for married men of 5 years, but barely one year for single men.
- Indicate that singles are driving the slowdown in female life expectancies.

		Married			Single		
		60	80	100	60	80	100
FEMALE	2001	23.94	9.49	0.80	21.48	8.66	0.80
	2010	24.53	10.21	0.85	21.72	8.42	0.81
	2020	25.23	11.05	0.89	21.95	8.16	0.82
	2030	25.94	11.93	0.93	22.14	7.90	0.83
	2040	26.67	12.80	0.95	22.30	7.64	0.84
	2050	27.37	13.66	0.97	22.44	7.39	0.85
MALE	2001	20.91	7.70	0.81	16.17	6.25	0.78
	2010	22.05	8.10	0.81	16.23	6.32	0.80
	2020	23.24	8.53	0.82	16.31	6.41	0.82
	2030	24.35	8.97	0.82	16.38	6.49	0.83
	2040	25.37	9.39	0.83	16.45	6.58	0.85
	2050	26.32	9.82	0.84	16.53	6.67	0.86

Region

.

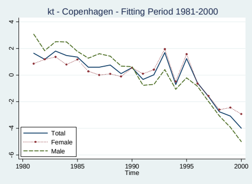

Beta - Copenhagen - Fitting Period 1981-2000
Beta - Greater Copenhagen - Fitting Period 1981-2000
Beta - Zealand & Falster - Fitting Period 1981-2000
Beta - Funen & Islands - Fitting Period 1981-2000
Beta - South Jutland - Fitting Period 1981-2000
Beta - West Jutland - Fitting Period 1981-2000
Beta - Central Jutland - Fitting Period 1981-2000
Beta - North Jutland - Fitting Period 1981-2000
Total
Female
Male
Age
60
70
80
90
100
-.1
-.05
0
.05
.1
.15

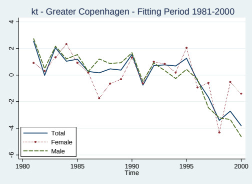
Region

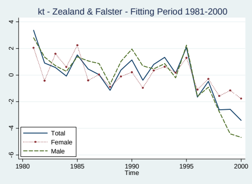
.

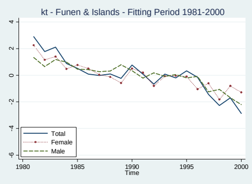
- We observe there are indeed differences in the estimates across regions.
- For males in all the regions we find a negative decline in the betas across ages except Zealand & Falster.
- For most females we find constant or negative betas up to age 80 and then the betas start increasing. Thus women have experienced a rise in (or constant) mortality at ages 60-80 compared to declining mortality at older ages.
- Thus for all the regions women's transition towards lower mortality has its foundation at ages 80-100, whereas for men it is centered around ages 60-80.

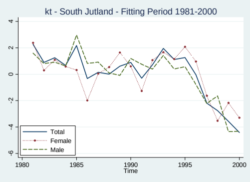
Region

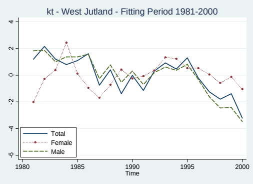

kt - Copenhagen - Fitting Period 1981-2000
Total
Female
Male
Time

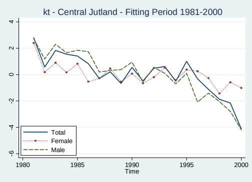

kt - Greater Copenhagen - Fitting Period 1981-2000
Total
Female
Male
Time

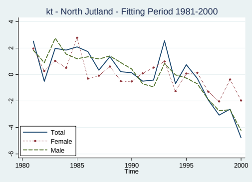

kt - Zealand & Falster - Fitting Period 1981-2000
Total
Female
Male
Time


kt - Funen & Islands - Fitting Period 1981-2000
Total
Female
Male
Time


kt - South Jutland - Fitting Period 1981-2000
Total
Female
Male
Time


kt - West Jutland - Fitting Period 1981-2000
Total
Female
Male
Time


kt - Central Jutland - Fitting Period 1981-2000
Total
Female
Male
Time


kt - North Jutland - Fitting Period 1981-2000
Total
Female
Male
Time

Region

- Life expectancies for the different regions using register data for all individuals aged 60-100 for the time period 1981-2000.

		Copenhagen			Greater Copenhagen			Zealand & Falster			Funen & Islands		
		60	80	100	60	80	100	60	80	100	60	80	100
F	2001	22.30	8.86	0.80	22.40	8.76	0.77	21.98	8.69	0.83	22.91	8.96	0.83
E	2010	22.61	9.14	0.81	22.54	8.92	0.79	22.14	9.03	0.85	23.22	9.47	0.83
M	2020	22.97	9.46	0.82	22.68	9.09	0.81	22.30	9.43	0.87	23.55	10.04	0.84
A	2030	23.33	9.78	0.83	22.83	9.26	0.83	22.47	9.85	0.89	23.85	10.60	0.84
L	2040	23.69	10.11	0.84	22.98	9.44	0.85	22.63	10.28	0.91	24.13	11.15	0.85
E	2050	24.06	10.44	0.84	23.12	9.62	0.86	22.79	10.72	0.92	24.38	11.69	0.85
	2001	18.59	7.02	0.77	19.49	6.89	0.77	18.76	6.96	0.78	19.44	6.97	0.73
M	2010	19.55	7.43	0.77	20.41	7.11	0.76	19.42	7.26	0.80	20.16	7.05	0.68
A	2020	20.60	7.89	0.77	21.37	7.34	0.75	20.14	7.61	0.83	20.91	7.14	0.63
L	2030	21.63	8.35	0.77	22.26	7.57	0.74	20.85	7.97	0.85	21.60	7.23	0.56
E	2040	22.63	8.80	0.76	23.08	7.79	0.72	21.56	8.35	0.87	22.23	7.31	0.50
	2050	23.60	9.25	0.76	23.83	8.01	0.71	22.27	8.74	0.89	22.81	7.39	0.43

		South Jutland			West Jutland			Central Jutland			North Jutland		
		60	80	100	60	80	100	60	80	100	60	80	100
F	2001	22.99	8.99	0.80	22.54	8.52	0.82	22.72	8.82	0.80	22.74	8.98	0.77
E	2010	23.43	9.26	0.82	22.57	8.55	0.83	22.92	9.20	0.82	22.95	9.34	0.80
M	2020	23.92	9.57	0.84	22.60	8.57	0.84	23.13	9.63	0.85	23.19	9.76	0.83
A	2030	24.40	9.89	0.86	22.63	8.60	0.85	23.34	10.06	0.87	23.44	10.20	0.85
L	2040	24.87	10.22	0.88	22.67	8.63	0.86	23.55	10.51	0.89	23.69	10.65	0.88
E	2050	25.34	10.56	0.89	22.70	8.66	0.87	23.75	10.97	0.91	23.94	11.11	0.89
	2001	19.69	6.89	0.77	19.67	6.89	0.77	19.75	7.08	0.78	19.74	6.92	0.80
M	2010	20.44	7.07	0.76	20.40	6.97	0.76	20.60	7.29	0.78	20.49	7.06	0.80
A	2020	21.22	7.26	0.75	21.14	7.06	0.75	21.48	7.52	0.77	21.27	7.21	0.79
L	2030	21.95	7.44	0.74	21.81	7.13	0.74	22.30	7.75	0.76	21.99	7.35	0.78
E	2040	22.63	7.63	0.73	22.42	7.21	0.73	23.07	7.97	0.76	22.65	7.50	0.77
	2050	23.26	7.80	0.72	22.97	7.28	0.72	23.77	8.19	0.75	23.25	7.64	0.76

Region

- In 2001, the forecasted life expectancies for females at age 60 exhibit a gap of one year with a low in Zealand & Falster of 22 and a high in South Jutland of 23 years.
- For males the spread is greater than a year, with a high in Central and North Jutland just below 20 and a low in Cph. of 18.5 years.
- Within regions, gender asymmetries appear as the gender gap is almost four years in Cph. but only 2.8 years in West Jutland.
- Over the forecast horizon of 50 years there are large variations in gains of life expectancy at age 60. The average increase is higher for males than females: Males gain 3.5-4 years across all regions, whereas females only gain 0.2-2.5 years.
- The highest gain in life expectancy is found in South Jutland and the lowest in Greater Copenhagen and Zealand & Falster.
- Furthermore, the forecasts in 2050 becomes higher for males than females in Greater Copenhagen and West Jutland.

Conclusion

- By developing more sophisticated data we are able to show that even when individuals share the same age and gender status, significant differences in life expectancies appear across marital status and regions.
- We find clear evidence that married individuals have lower mortality rates compared to singles and this pattern will continue in the future. This confirms that marriage in itself indeed is beneficial in terms of life expectancies.
- The regional specific forecast displays smaller differences in future life expectancies. However still very relevant as differences in years between regions are observed and huge gender differences exist.
- Care should be taken as the mortality forecasts are sensitive to the choice of calibration period.

New Sub-populations

- At the moment we are in the process of adding further sub-populations.
- Education: It is no possible to create mortality forecasts based on level of education as not enough individuals within each age have a high enough education. Remember these individuals where born in the first part of the 20th century where relatively few people got a higher education.
- Income: As people belonging to more upmarket socio-economic groups tend to live longer the idea is to use income as a proxy. Currently in the process of calculating these forecasts.
- Trying to come up with ideas as to how we can exploit the extensive health data.

New Model for Mortality Forecasting

- Joint work with Bent Nielsen at Oxford University.
- We develop a potentially powerful new method for mortality forecasting that allows for explanatory variables:

$$\ln m_x(t) = \alpha_x + \beta_t + \gamma_{t-x} + f(y_{x,t})$$

- To ensure a unique set of parameter estimates - thereby avoiding the classic identification problem - we use a canonical parameterisation as in Kuang-Nielsen-Nielsen (2008)
- The data will be aggregated but derived from an individual-based panel of the population accounting for population characteristics such as marital status, income, health, region, and trends in alcohol and smoking related diseases.