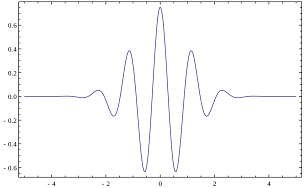
WAVELET-BASED FEATURE-ENGINEERING FOR MORTALITY PROJECTION

D. Hainaut[1], M. Denuit[1]

[1] UCLouvain - ISBA

Paris Actuarial Colloquium 2020

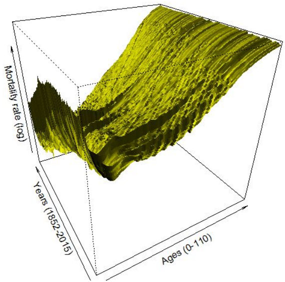
Motivation



Example: Daubechies wavelet function

- Wavelet theory is a powerful tool for:
  - compressing time series or images
  - denoising time series or images
- Wavelet transform
  - projects a signal on an orthonormal basis of functions
  - provides a sparse representation of the data.

Motivation



Belgian Log-Mortality rates (1852-2015)

Research questions:

1. Can wavelets denoise / smooth curves of log-mortality rates?
2. Does the wavelet transform provide a sparse representation of the surface of log-mortality?
3. Can we use wavelets for forecasting future mortality?

Wavelets in a nutshell

- Wavelets are defined by parent functions: a "father" wavelet ϕ and a "mother" wavelet ψ orthogonal to ϕ (i.e. $\int_{\mathbb{R}} \phi(x)\psi(x)dx = 0$)
- The wavelet basis is obtained by dilating and translating ψ to form the dictionary $\mathcal{D}$, that is, $\psi(\frac{x-b}{a})$ for $a > 0$ and $b \in \mathbb{R}$.
- To ensure sparsity: $a = 2^j$ and $b = k2^{-j}$ where k and j are integers.
- The functions ϕ and

$$\psi_{j,k}(x) = 2^{j/2}\psi\left(2^j x - k\right) \text{ for } k, j \in \mathbb{Z}$$

form the available dictionary $\mathcal{D}$ to model the mortality curve.

Wavelets in a nutshell

- If $n_{t,x}$ is the number of deaths and $E_{t,x}$ is the exposure, an unbiased estimator $\widehat{\mu}(t,x)$ of the force of mortality $\mu(t,x)$ (year t, age x) is

$$\widehat{\mu}(t,x) = \frac{n_{t,x}}{E_{t,x}}.$$

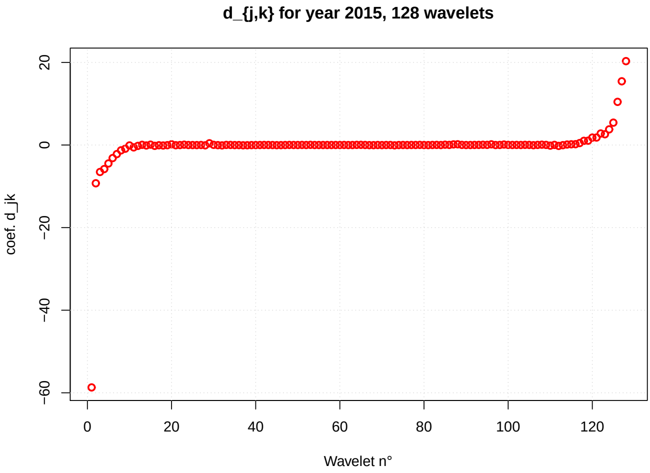
- The discrete wavelet transform (DWT) decomposes $\ln \widehat{\mu}(t,x)$ as a sum of wavelet functions, for a fixed calendar year t

$$\ln \widehat{\mu}(t,x) = c_0(t)\phi(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}(t)\psi_{j,k}(x).$$

- The vector of wavelet coefficients $\boldsymbol{d}(t) = \left(d_{j,k}(t)\right)_{j,k}$ is of dimension 2^J (e.g. $2^J = 128$).

DWT applied to Belgian mortality (year 2015)

- For a given year (here 2015) Most of $d_{j,k}$ are close to zero... We can then "trash" many of them! But how many?



A chi-square test for thresholding $d_{j,k}(t)$

- Donoho and Johnstone (1994) propose hard thresholding: we cancel all wavelet coefficients smaller in absolute value than a threshold d^*:

$$\widehat{d}_{j,k}(t) = \begin{cases} 0 \text{ if } |d_{j,k}(t)| < d^* \\ d_{j,k}(t) \text{ otherwise} \end{cases}$$

- $\widehat{\mu}^S(t,x)$ the shrinked mortality rates are constructed as:

$$\ln \widehat{\mu}^S(t,x) = c_0(t)\phi(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \widehat{d}_{j,k}(t)\psi_{j,k}(x) \,. \qquad (1)$$

How do we choose d^*?

A chi-square test for thresholding $d_{j,k}(t)$

If the Cochran (1952) criterion ($n_{t,x}\widehat{q}_{t,x} \geq 5$) is fulfilled and the population is large and Normal approximation holds:

$$\widehat{\mu}(t,x) \sim \text{Normal}\left(\mu(t,x)\,,\ \frac{\mu(t,x)}{E_{t,x}}\right)\,. \tag{2}$$

We test for:

$$\begin{cases} H_0 \ : \quad \mu(t,x) = \widehat{\mu}^S(t,x)\,, \\ H_1 \ : \quad \mu(t,x) \neq \widehat{\mu}^S(t,x)\,, \end{cases}$$

with the Chi-Square statistics

$$S_t = \sum_{x=x_{min}}^{x_{max}} E_{t,x} \frac{\left(\widehat{\mu}^S(t,x) - \widehat{\mu}(t,x)\right)^2}{\widehat{\mu}^S(t,x)}\,.$$

χ^2 with $n-p-1$ df (n number of observations, p number of non-null wavelet coefficients).

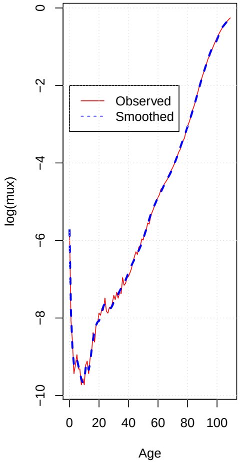
χ^2 selection, Belgian mortality year 2015

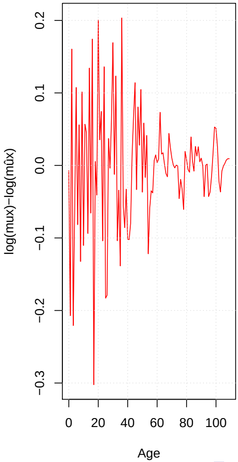
If S_t is too large, rejection of H_0. For the year 2015: the curve $\widehat{\mu}^S(t,x)$ with 22 Daubechies wavelets (order 4) is not rejected. For higher threshold: rejection of $\mu(t,x) = \widehat{\mu}^S(t,x)$.

Threshold d^*	$p = \#$ of $\widehat{d}_{j,k} \neq 0$	S_t	χ^2 97.5%	AIC	BIC
0.01	93	10.08	16.01	-1328.73	-1077.58
0.03	60	24.74	59.34	-1380.04	-1218.01
0.05	49	38.99	72.62	-1387.8	-1255.47
0.11	35	60.24	89.18	-1394.52	-1300
0.15	30	78.19	95.02	-1386.07	-1305.05
0.19	26	87.56	99.68	-1384.76	-1314.54
0.21	24	96.3	102	-1380.05	-1315.23
0.23	22	98.1	104.32	-1382.24	-1322.83

χ^2 selection, Belgian mortality year 2015

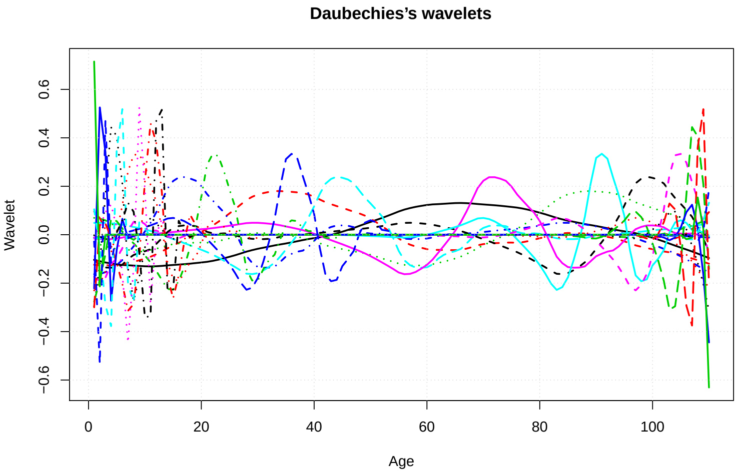
Hard shrinkage of wavelets coefficients also smoothes the curve of log-mortality rates!





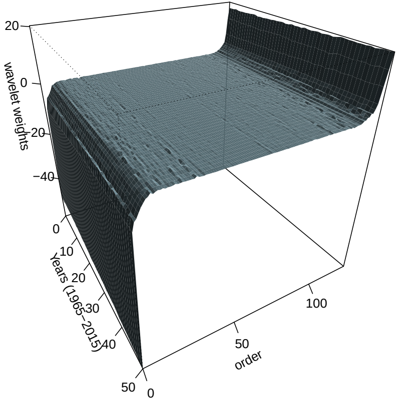
χ^2 selection, Belgian mortality year 2015

The 22 Daubechies wavelets (order 4) needed to approximate the 2015 mortality rates.



DWT applied to Belgian mortality surface (1965-2015)

Remarkable observations: wavelets with null coefficients are the same for every calendar year! A small number of wavelets is needed for smoothing all mortality curves between 1965 and 2015.



Multi-years wavelets selection

- The selection of wavelets explaining the mortality over 1965 - 2015 is done by thresholding based on the average of wavelet coefficients.
- If T and $d^*$are the number of years in the data set and the threshold. If

$$\frac{1}{T}\left|\sum_{t=1}^{T} d_{j,k}(t)\right| < d^*$$

then we set $\widehat{d}_{j,k}(t) = 0$ for all $t \in \{1, ..., T\}$.

- We use as criterion for d^*, the AIC and BIC :

$$\mathrm{AIC} = 2\,(T\,p) - 2\ln \mathcal{L}_{Poiss}(d^*, \boldsymbol{\mu}^S)\,,$$
$$\mathrm{BIC} = \ln(T\,n)\,(T\,p) - 2\ln \mathcal{L}_{Poiss}(d^*, \boldsymbol{\mu}^S)\,.$$

where $\ln \mathcal{L}_{Pois}(d^*, \boldsymbol{\mu}^S)$ is the log-likelihood summed up over all years.

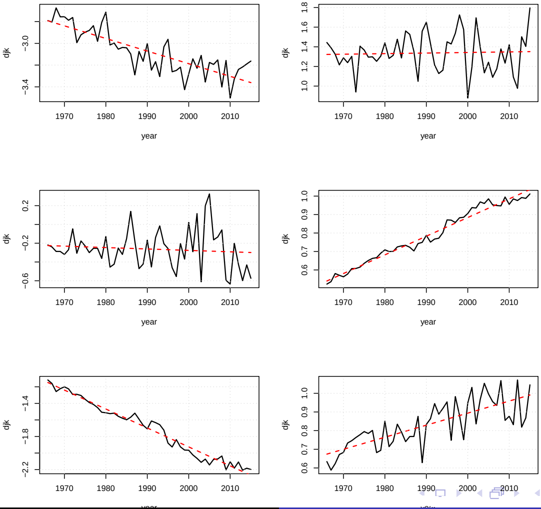
Multi-years wavelets selection

The lowest BIC is achieved with 24 Daubechies wavelets (Belgian mortality 1965-2015)

d^*	$p = \#$ of $\widehat{d}_{j,k} \neq 0$	% of χ^2 passed	$\ln \mathcal{L}(d^*, \mu^S)$	AIC	BIC
0.03	33	0.76	-23103.12	49572.24	60734.41
0.06	30	0.76	-23167.18	49394.35	59541.78
0.09	28	0.76	-23221.25	49298.5	58769.44
0.12	24	0.59	-23785.23	50018.46	58136.4
0.15	24	0.59	-23785.23	50018.46	58136.4
0.18	24	0.59	-23785.23	50018.46	58136.4
0.21	24	0.59	-23785.23	50018.46	58136.4
0.24	24	0.59	-23785.23	50018.46	58136.4
0.27	22	0.31	-24318.95	50881.89	58323.34
0.3	21	0	-29637.46	61416.93	68520.13

Multi-years wavelets selection

The time series of 24 $d_{j,k}(t)$ for or $t = 1965$ to 2015, display linear trends! Here 6 examples:



Multi-years wavelets selection

- Based on this remarkable observation we regress linearly coefficients with respect to time. For $\widehat{d}_{j,k}(t) \in \widehat{\boldsymbol{d}}(t)$ we fit:

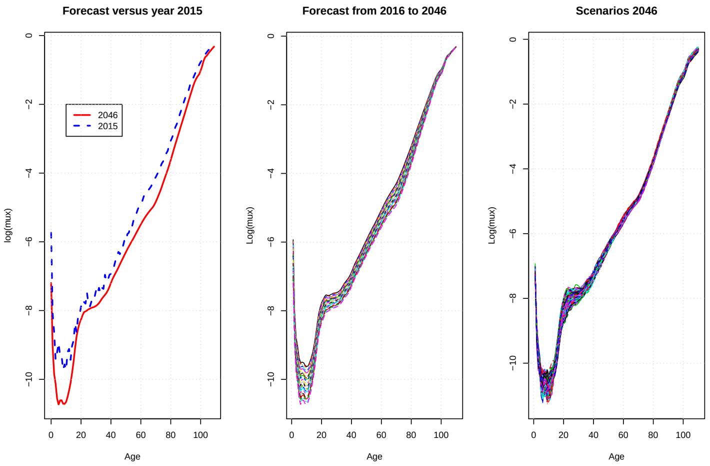
$$\widehat{\boldsymbol{D}}(t) = \boldsymbol{\alpha} + \boldsymbol{\beta}\, t + \boldsymbol{\mathcal{E}}(t)$$

  where $\boldsymbol{\mathcal{E}}(t)$ are independent, multivariate Normal random vectors with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}$.

- 62.5% of linear regressions have a R^2 above 75%
- Assumption of Normality (at 5%) for residuals is satisfied for 22 time series out of 24.
- By linear extrapolation, we simulate future wavelet coefficients and mortality rates.

Forecasting

Belgian population. Left plot: $\ln \mu(t, x)$ in 2015 and 2046. Mid plot: Average $\ln \mu(t, x)$ from 2016 to 2046 (1000 sim.). Right plot: 1000 sim. of $\ln \mu(t, x)$, year 2046.



Belgian population: Evolution of forecast cross-sectional life expectancies from 2016 to 2045 (24 Daubechies wavelets, order 4).

	Cross sectional life expectancy				
Year	At birth	Age 20	Age 40	Age 60	Age 80
2016	80.61	61.03	41.69	23.58	8.62
2020	81.34	61.7	42.33	24.14	8.9
2025	82.2	62.51	43.1	24.81	9.25
2030	83.05	63.31	43.86	25.47	9.61
2035	83.84	64.06	44.58	26.1	9.94
2040	84.62	64.81	45.3	26.72	10.29
2045	85.37	65.53	45.99	27.33	10.63

Validation by back testing

- Benchmark models:

Lee Carter 1992	$\ln \mu(t,x) = \alpha_x + \beta_x \kappa_t$
Renshaw et al. 2003	$\ln \mu(t,x) = \alpha_x + \beta_x^1 \kappa_t^1 + \beta_x^2 \kappa_t^2,$
Renshaw et al. 2006	$\ln \mu(t,x) = \alpha_x + \beta_x^1 \kappa_t + \beta^2 \gamma_{t-x}$
Cairns et al. 2006	$\text{logit}\, q(t,x) = \kappa_t^{(1)} + (x - \bar{x})\, \kappa_t^{(2)}$
Cairns et al. 2009	$\text{logit}\, q(t,x) = \kappa_t^{(1)} + (x - \bar{x})\, \kappa_t^{(2)}$

- The six models are fitted to Belgian data from 1965 to 2005 for ages ranging from 0 to 90 years.

- Next we forecast mortality for years 2006 to 2015 and compare with observed ones.

- Sum of squared errors between forecast and real log-mortality rates

	Wavelet (24 coef.)	LC	RH 2D	RH coh	CBD	M7
2006	2.11	2.59	1.96	4.32	64.57	41.83
2007	1.14	1.43	1.37	12.51	67.44	48.87
2008	1.29	1.62	1.21	29.54	65.32	56.02
2009	1.77	2.77	2.59	53.47	67.54	59.41
2010	4.01	5.12	4.91	87.54	69.31	76.11
2011	3.01	3.77	4.03	124.45	72.75	71.38
2012	5.18	6.3	6.4	179.46	74.49	78.79
2013	3.81	5.42	4.86	235.79	71.57	97.82
2014	4.00	5.49	5.48	294.46	70.09	105.42
2015	4.69	5.95	5.41	380.99	71.2	107.2

Validation by back testing

- Population: USA. Sum of squared errors between forecast and real log-mortality rates

	Wavelet (24 coef.)	LC	RH 2D	RH coh	CBD	M7
2006	0.26	0.52	0.46	36.52	43.62	48.69
2007	0.25	0.54	0.46	65.87	44.73	57.31
2008	0.29	0.71	0.51	107.93	44.57	70.68
2009	0.54	0.93	0.7	162.91	44.46	81.73
2010	0.62	1.1	0.79	233.54	44.47	97.75
2011	0.6	1.05	0.83	320.87	46.39	109.47
2012	0.68	1.14	0.96	427.44	47.85	124.44
2013	0.8	1.2	1.1	553.83	49.36	136.64
2014	1.03	1.23	1.23	694.87	51.32	151.17
2015	1.44	1.32	1.64	861.92	57.15	160.82

Validation by back testing

- Population: UK. Sum of squared errors between forecast and real log-mortality rates

	Wavelet (24 coef.)	LC	RH 2D	RH coh	CBD	M7
2006	0.45	1.58	0.64	0.53	74.86	42.67
2007	0.37	1.53	0.58	0.55	76.79	49.9
2008	0.76	2.18	0.88	0.79	78.88	58.97
2009	0.63	1.61	0.79	0.65	76.38	72.51
2010	1.14	2.17	1.31	0.85	76.56	80.78
2011	1.78	2.5	1.95	0.96	74.75	97.33
2012	2.15	2.98	2.36	1.74	75.64	103.88
2013	1.76	2.77	2.19	2.19	77.16	118.45
2014	1.8	3.12	2.21	2.98	82.1	124.52
2015	1.52	3	1.95	4.68	85.23	138.8

Comparison of life expectancies Belgium, USA, UK

- Belgian, US and UK populations: Evolution of forecast cross-sectional life expectancies at birth from 2016 to 2045 (Wavelet model, 24 coefficients).

Year	Belgium	USA	UK
2016	80.61	79.01	80.47
2020	81.34	79.61	81.16
2025	82.2	80.32	81.96
2030	83.05	81.01	82.73
2035	83.84	81.68	83.46
2040	84.62	82.32	84.21
2045	85.37	82.95	84.87

Conclusions

- Wavelets are powerful tools for analyzing mortality trends.
- The chi-square test allows to smooth log-mortality rates by wavelets shrinkage.
- With both approaches : 110 death rates summarized by twenty wavelet coefficients.
- A small number of wavelets can reconstruct all curves of mortality between 1965 and 2015.
- Coefficients exhibit clear trends easy to extrapolate with a basic multivariate linear regression.
- Wavelet model widely outperforms other popular actuarial models fitted to Belgian, US and UK populations, both in terms of predictive errors.