

Using neural networks in a Life & Health insurance context

SAV après-midi, 15 September 2022

Dr. Daniel Meier, Aktuar SAV
Swiss Re Institute



Key take-aways

- Neural networks should be part of the model toolbox of Life & Health actuaries
- There are plenty of model libraries, tutorials, online courses available – and data availability is also improving
- **Try it out...**



3 examples

Agenda

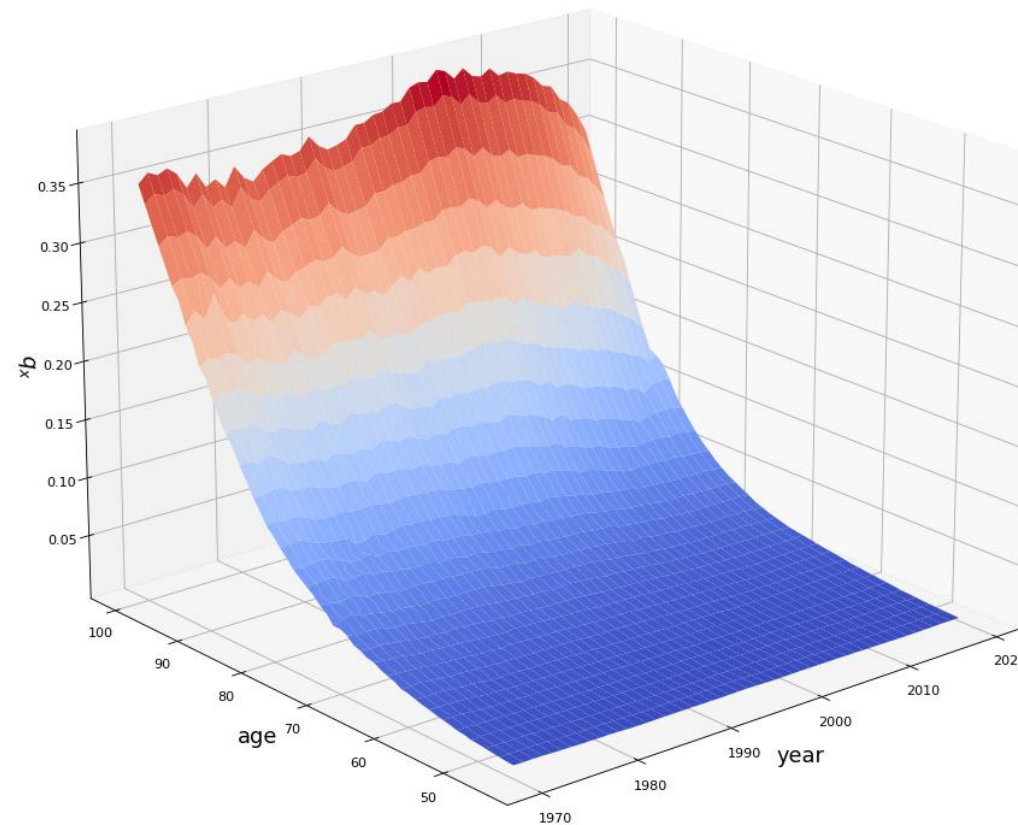
1. Mortality forecasting
2. Detecting anomalies in mortality rates
3. Neural networks as an alternative to classical survival models

Swiss Re
Swiss Re Institute

Swiss Actuarial Association
Data Science Working Group
actuarialdatascience.org

Example 1: Mortality forecasting

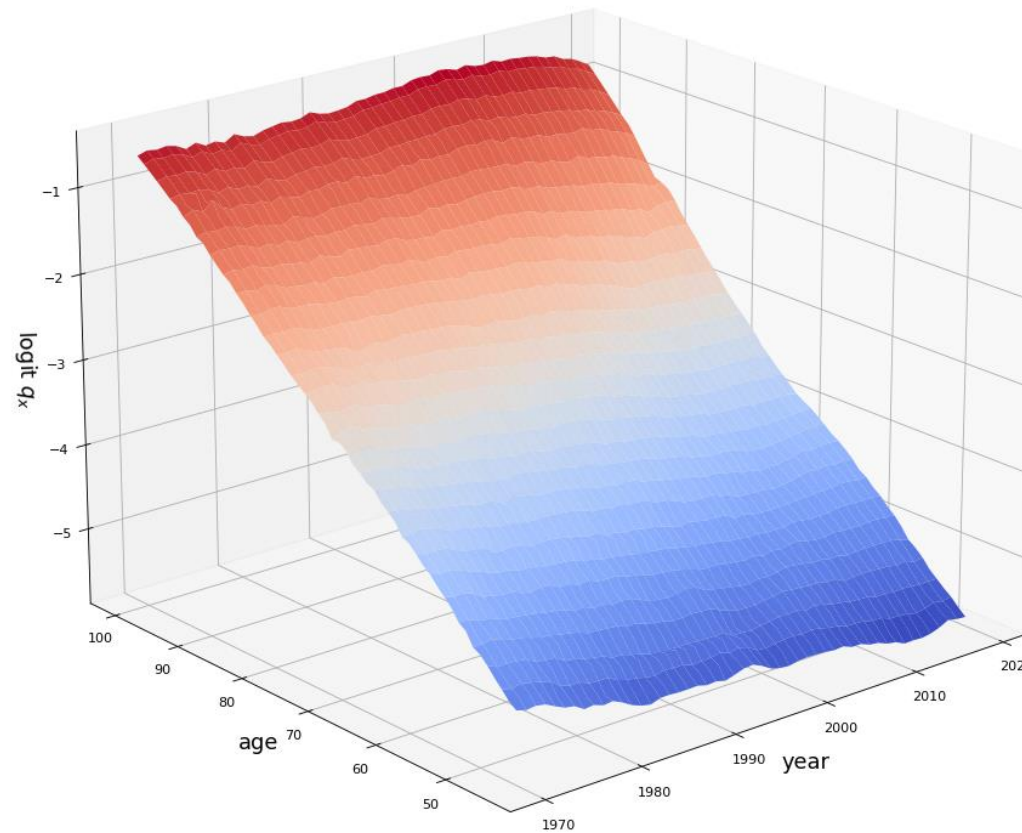
Mortality rates $q_{x,t}$ US males, 1970 to 2019



Data source: mortality.org

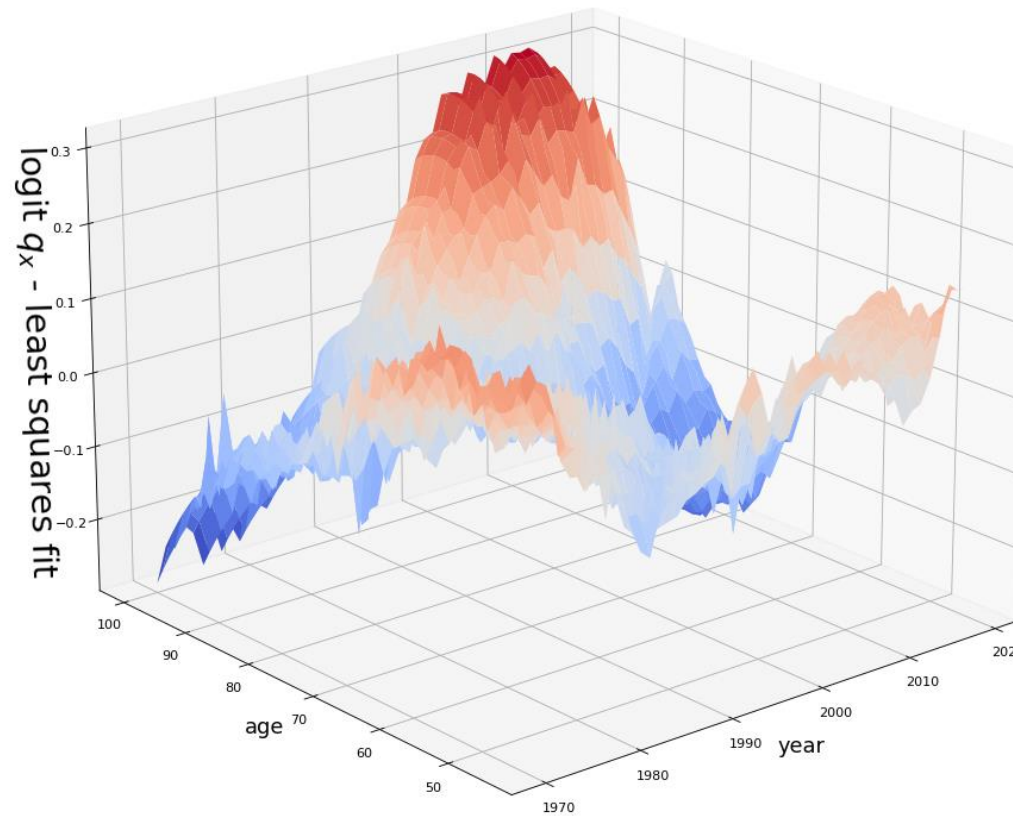
Example 1: Mortality forecasting

logit $q_{x,t}$ US males, 1970 to 2019



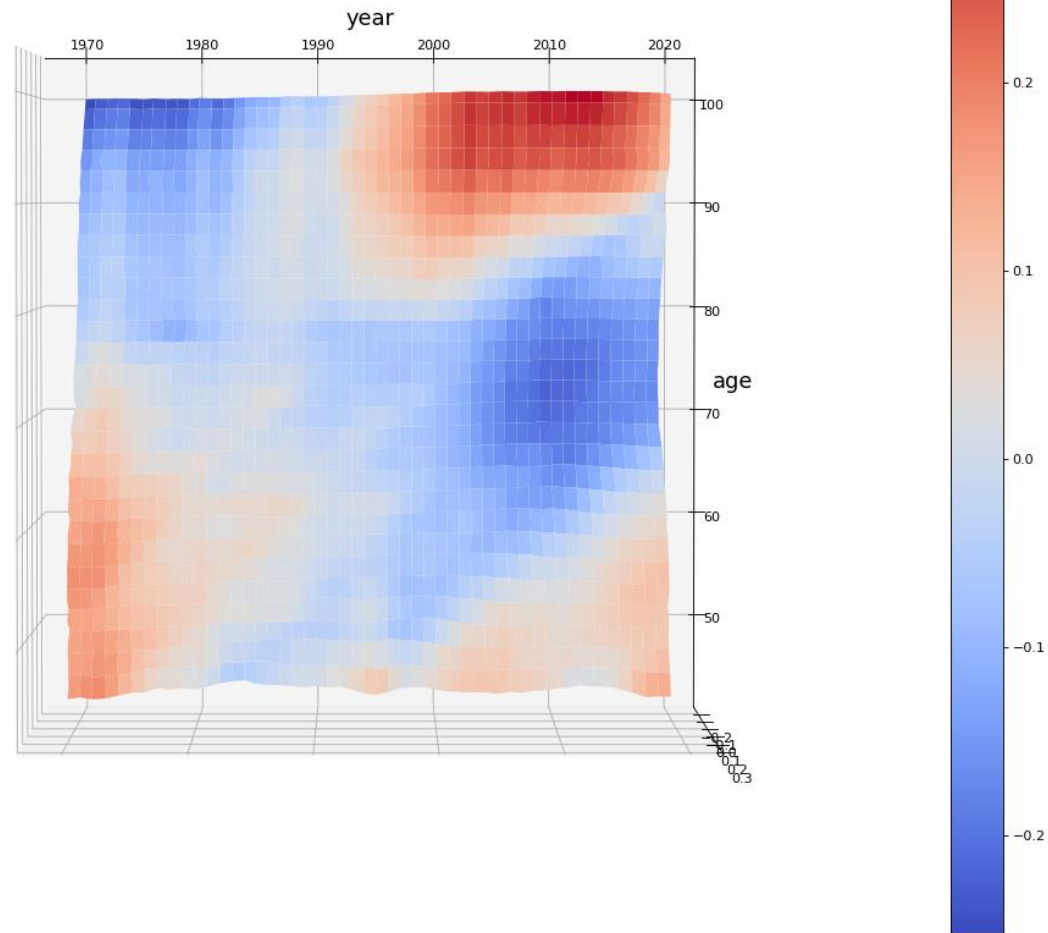
Example 1: Mortality forecasting

$\text{logit } q_{x,t} - (a_0 + a_1x + a_2t)$ based on
least squares US males, 1970 to 2019



Example 1: Mortality forecasting

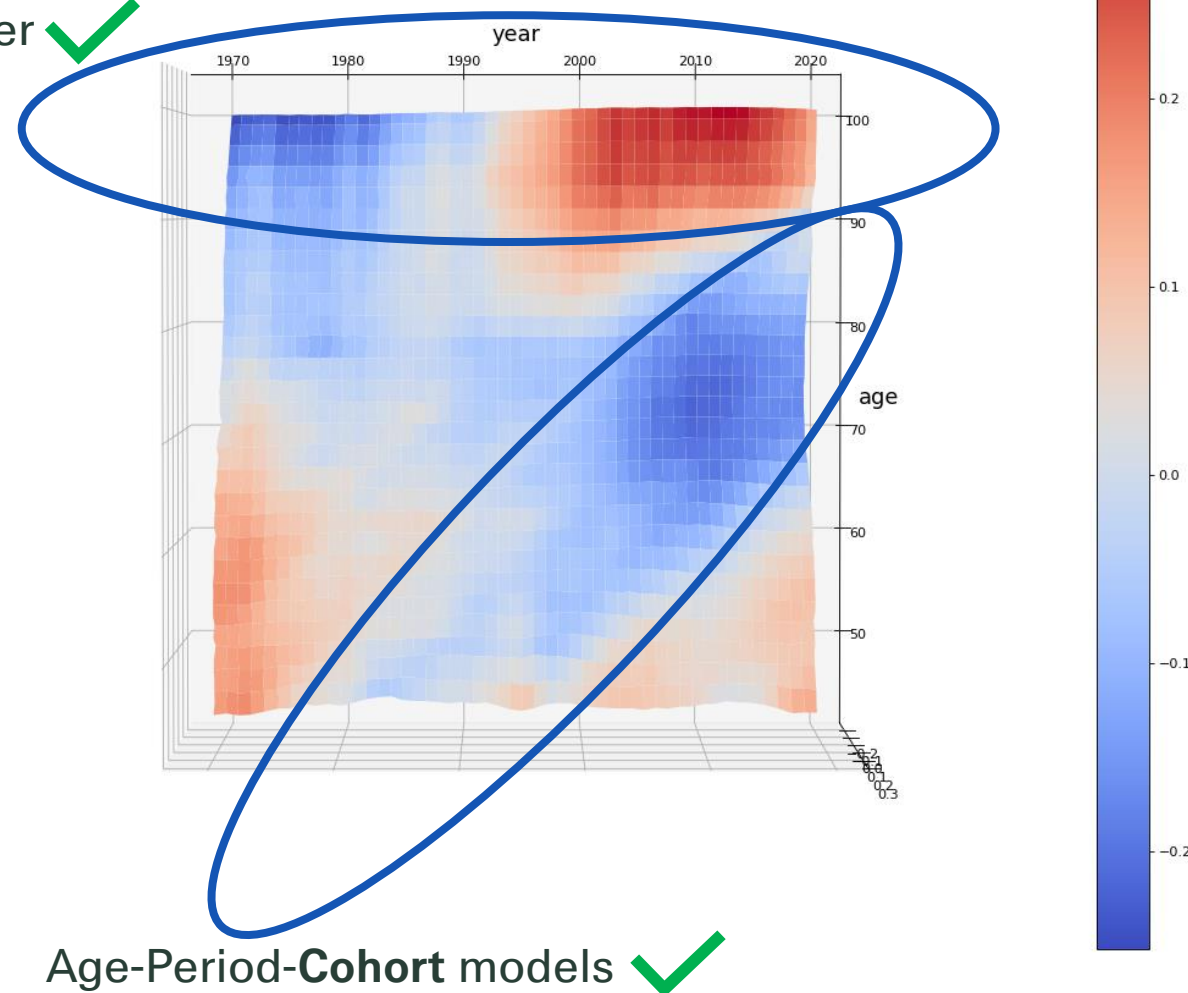
$\text{logit } q_{x,t} - (a_0 + a_1x + a_2t)$ based on
least squares US males, 1970 to 2019



Example 1: Mortality forecasting

$\text{logit } q_{x,t} - (a_0 + a_1x + a_2t)$ based on
least squares US males, 1970 to 2019

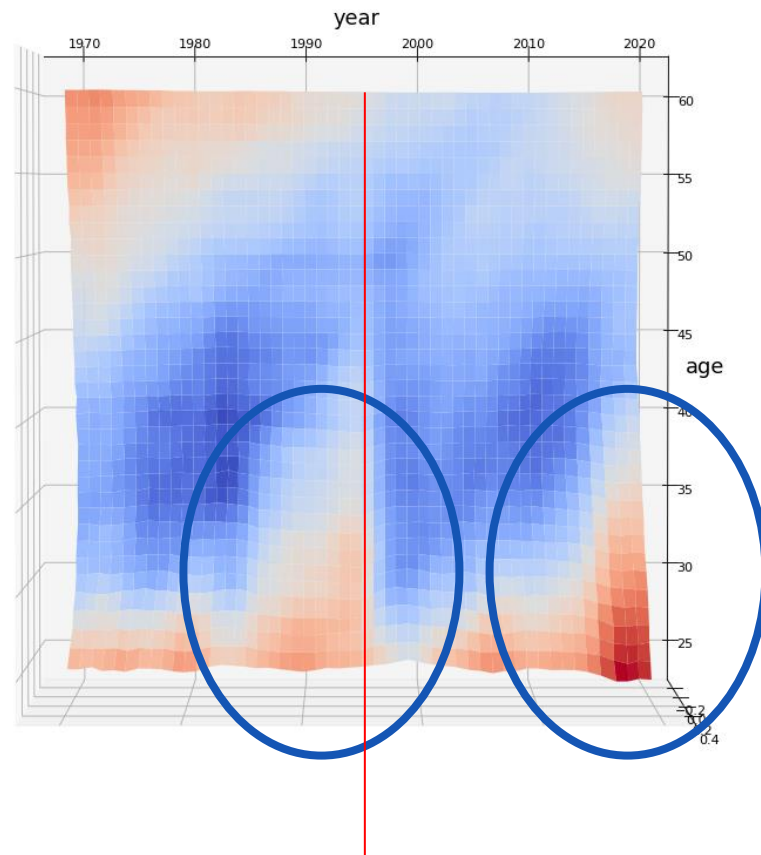
Lee-Carter ✓



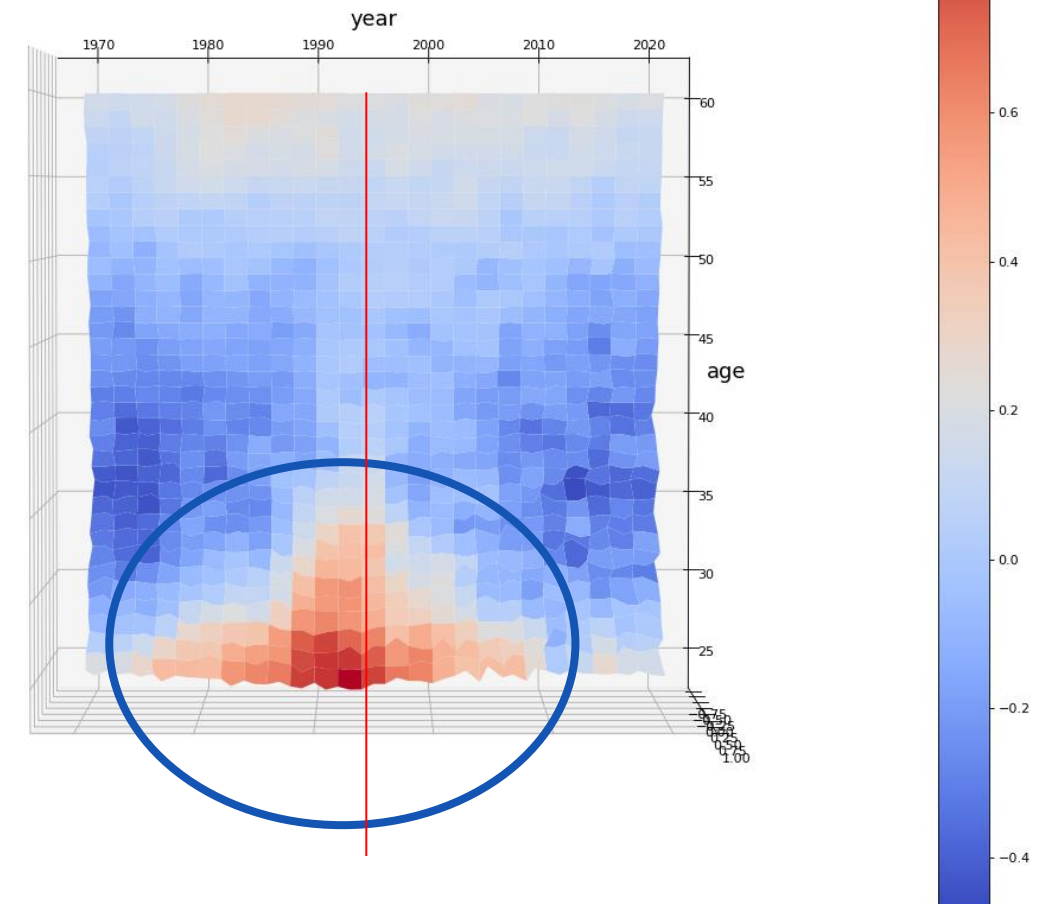
Age-Period-Cohort models ✓

Example 1: Mortality forecasting

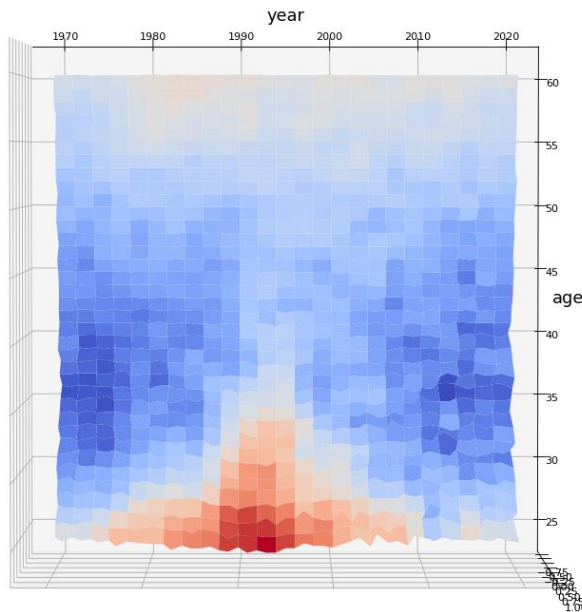
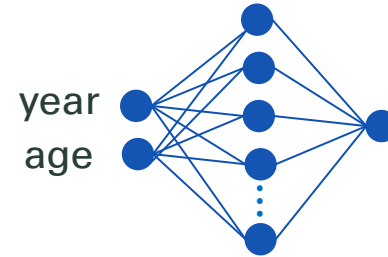
$\text{logit } q_{x,t} - (a_0 + a_1x + a_2t)$ based on
least squares US males, 1970 to 2019



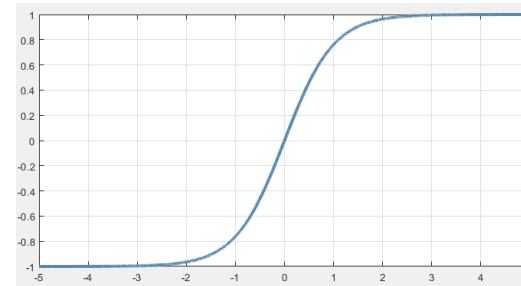
$\text{logit } q_{x,t} - (a_0 + a_1x + a_2t)$ based on
least squares CH males, 1970 to 2019



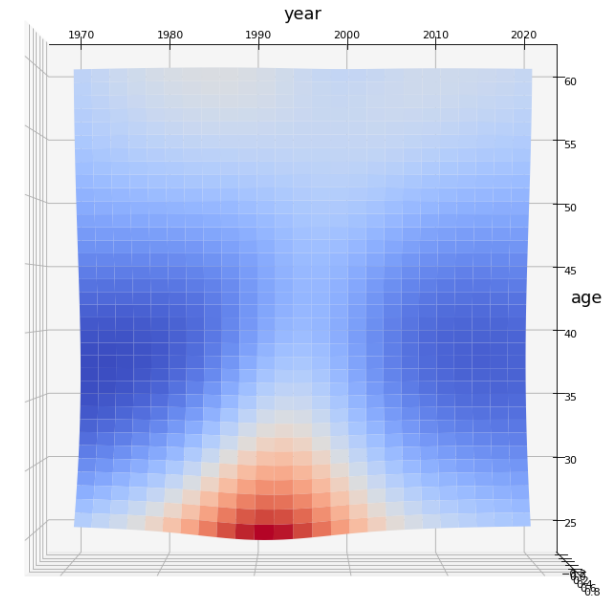
Example 1: Mortality forecasting



```
from tensorflow import keras  
model = Sequential()  
model.add(Dense(100))  
model.add(Activation('tanh'))  
model.add(Dense(1))  
model.compile(loss=keras.losses.mean_squared_error)  
summary = model.fit(x=X, y=Y, epochs=800, verbose=0, validation_split=0.1)  
fit = model.predict(X)
```



Model fitting (no forecasting yet)



Forecasting: See for example case study 6 at actuarialdatascience.org

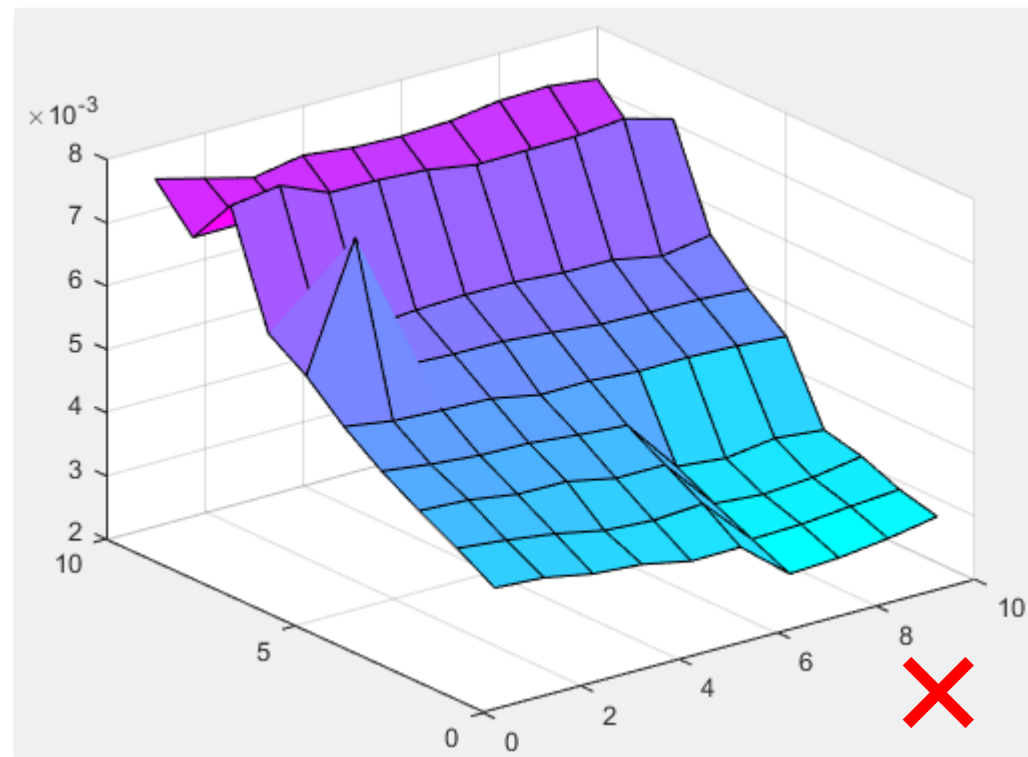
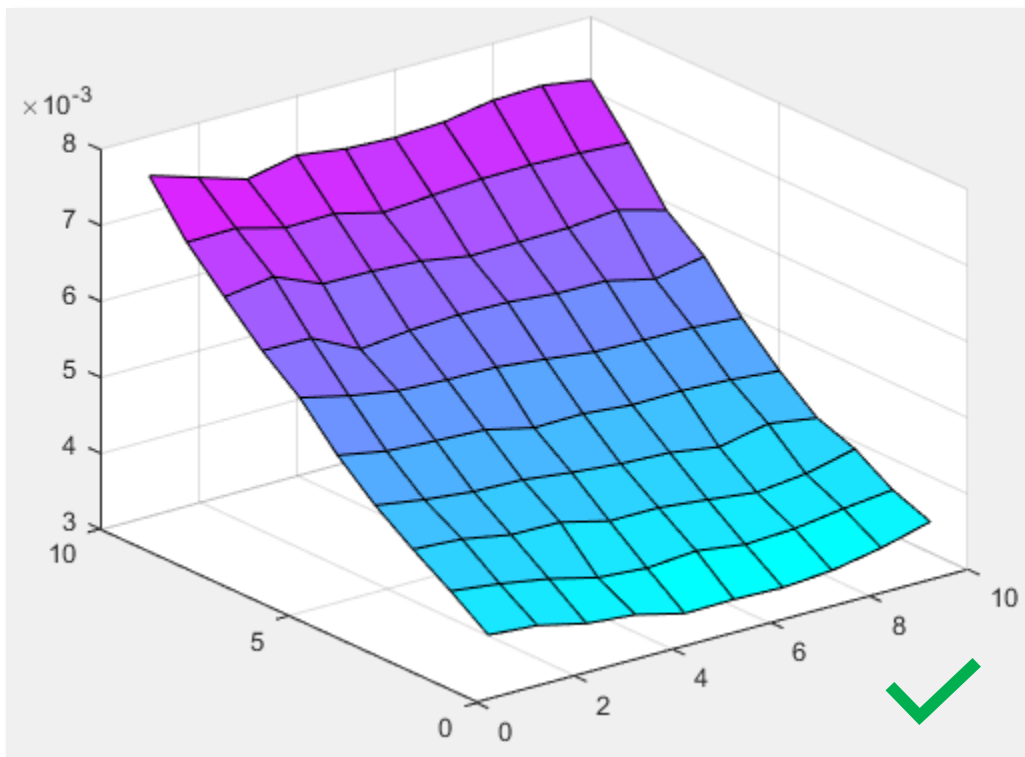
Example 2: Detecting anomalies in mortality rates

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
46	0.00349	0.00344	0.00329	0.00323	0.00307	0.00308	0.00306	0.00311	0.00323	0.0034
47	0.00384	0.00376	0.00365	0.00349	0.00344	0.00349	0.0034	0.00343	0.00352	0.00359
48	0.00416	0.00412	0.004	0.00399	0.00383	0.00383	0.00376	0.00368	0.00375	0.0039
49	0.00451	0.00441	0.00434	0.00432	0.00424	0.00418	0.00415	0.00405	0.00418	0.00408
50	0.00494	0.00482	0.00478	0.00475	0.0046	0.00461	0.00455	0.00456	0.00453	0.00447
51	0.00547	0.00532	0.00521	0.00518	0.00517	0.0051	0.00502	0.005	0.00499	0.00494
52	0.00586	0.00585	0.00553	0.0056	0.00563	0.00562	0.00556	0.00555	0.0054	0.00552
53	0.00634	0.00643	0.00616	0.00615	0.00611	0.006	0.00601	0.00602	0.00608	0.0059
54	0.00683	0.00684	0.00667	0.00668	0.00652	0.00657	0.00661	0.00662	0.0066	0.00656
55	0.00748	0.00728	0.00708	0.00722	0.00713	0.0071	0.00713	0.00724	0.00726	0.00716

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
46	0.00349	0.00344	0.00329	0.00323	0.00307	0.00308	0.00245	0.00249	0.00258	0.00272
47	0.00384	0.00376	0.00365	0.00349	0.00344	0.00349	0.00272	0.00274	0.00282	0.00287
48	0.00416	0.00412	0.004	0.00399	0.00383	0.00383	0.00301	0.00294	0.003	0.00312
49	0.00451	0.00441	0.00434	0.00432	0.00424	0.00418	0.00332	0.00324	0.00334	0.00326
50	0.00494	0.00482	0.00478	0.00475	0.0046	0.00461	0.00455	0.00456	0.00453	0.00447
51	0.00547	0.00745	0.00521	0.00518	0.00517	0.0051	0.00502	0.005	0.00499	0.00494
52	0.00586	0.00585	0.00553	0.0056	0.00563	0.00562	0.00556	0.00555	0.0054	0.00552
53	0.00761	0.00772	0.00739	0.00738	0.00733	0.0072	0.00721	0.00722	0.0073	0.00708
54	0.00683	0.00684	0.00667	0.00668	0.00652	0.00657	0.00661	0.00662	0.0066	0.00656
55	0.00748	0.00728	0.00708	0.00722	0.00713	0.0071	0.00713	0.00724	0.00726	0.00716

Which of the two subsets of mortality rates have anomalies?

Example 2: Detecting anomalies in mortality rates



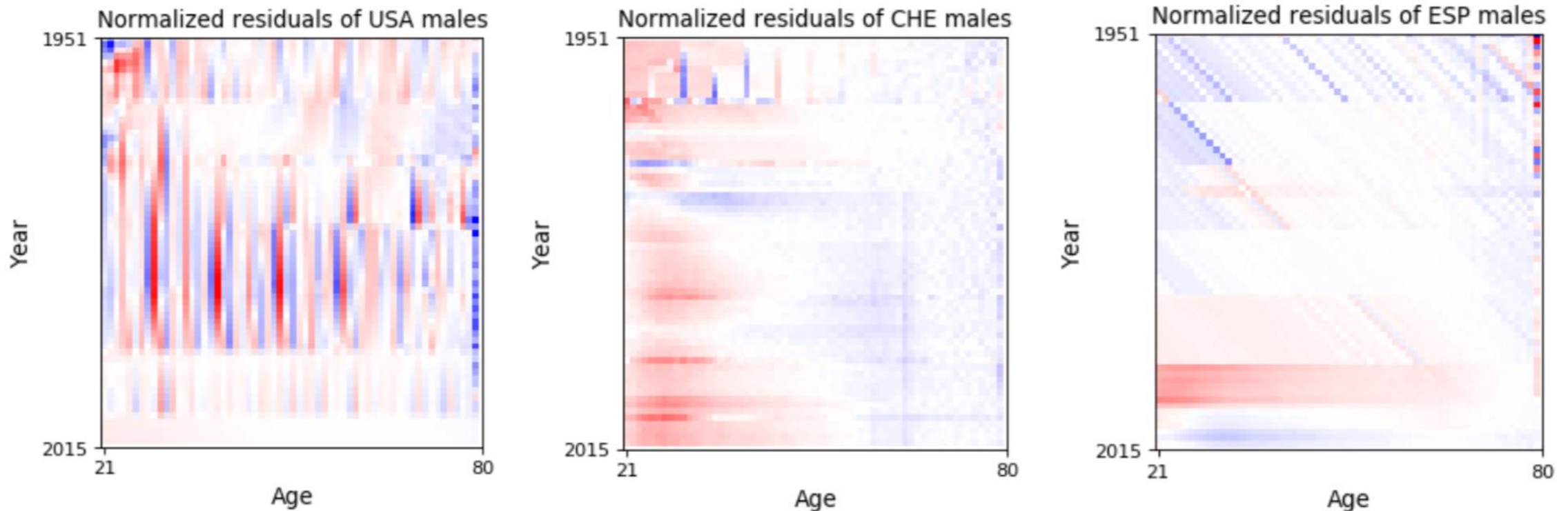
Which of the two subsets of mortality rates have anomalies?

Example 2: Detecting anomalies in mortality rates

Constructing training data: $E_{x+1,t+1} = E_{x,t}(1 - q_{x,t})$

Example 2: Detecting anomalies in mortality rates

Constructing training data: $E_{x+1,t+1} = E_{x,t}(1 - q_{x,t})$

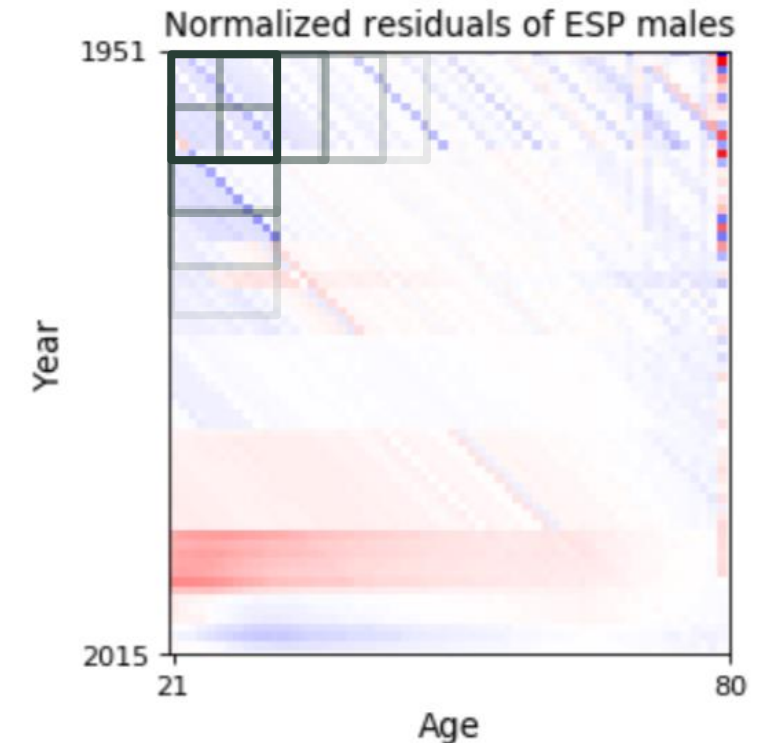
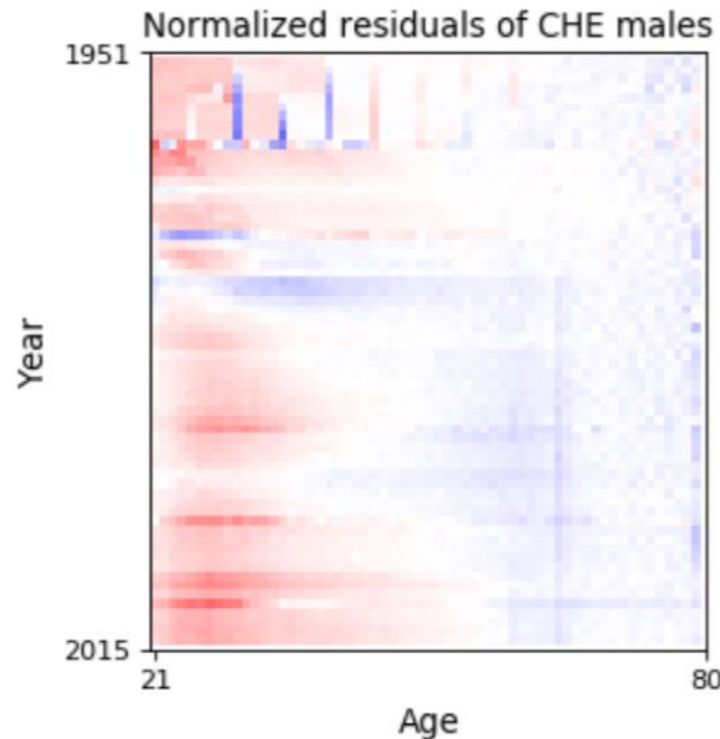
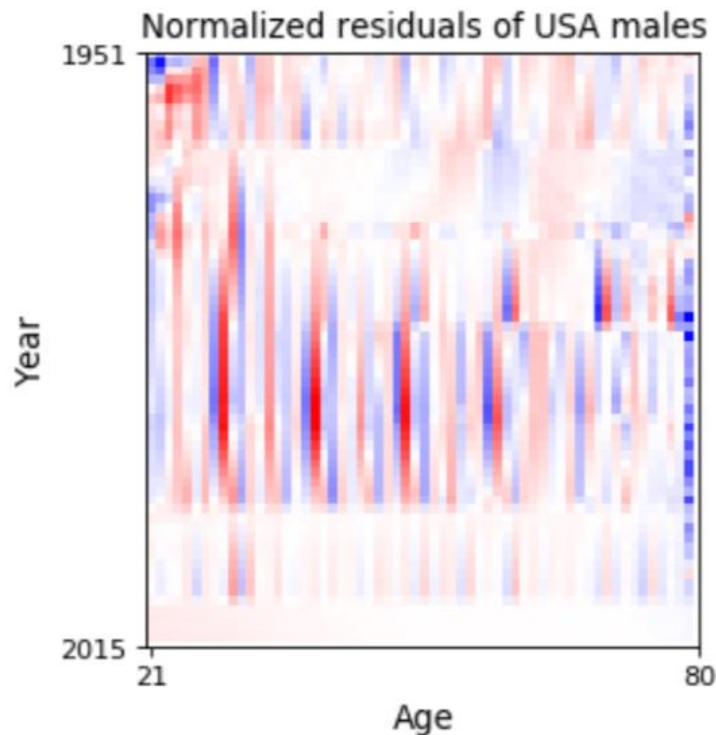


5 year census, groupings, methodologies, errors, migration

Example 2: Detecting anomalies in mortality rates

Constructing training data: $E_{x+1,t+1} = E_{x,t}(1 - q_{x,t})$

Model input: 10x10 mortality rates
Model output/labels: max abs residual



5 year census, groupings, methodologies, errors, migration

Example 2: Detecting anomalies in mortality rates

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
46	0.00349	0.00344	0.00329	0.00323	0.00307	0.00308	0.00306	0.00311	0.00323	0.0034
47	0.00384	0.00376	0.00365	0.00349	0.00344	0.00349	0.0034	0.00343	0.00352	0.00359
48	0.00416	0.00412	0.004	0.00399	0.00383	0.00383	0.00376	0.00368	0.00375	0.0039
49	0.00451	0.00441	0.00434	0.00432	0.00424	0.00418	0.00415	0.00405	0.00405	0.00405
50	0.00494	0.00482	0.00478	0.00475	0.0046	0.00461	0.0045	0.0045	0.00456	0.00456
51	0.00547	0.00532	0.00521	0.00518	0.00517	0.0051	0.00502	0.005	0.00499	0.00494
52	0.00586	0.00585	0.00553	0.0056	0.00563	0.00562	0.00556	0.00555	0.00555	0.00552
53	0.00634	0.00643	0.00616	0.00615	0.00611	0.006	0.00601	0.00602	0.00602	0.00602
54	0.00683	0.00684	0.00667	0.00668	0.00652	0.00657	0.00661	0.00662	0.00662	0.00662
55	0.00748	0.00728	0.00708	0.00722	0.00713	0.0071	0.00713	0.00724	0.00724	0.00724

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
46	0.00349	0.00344	0.00329	0.00323	0.00307	0.00308	0.00245	0.00249	0.00258	0.00272
47	0.00384	0.00376	0.00365	0.00349	0.00344	0.00349	0.00272	0.00274	0.00282	0.00287
48	0.00416	0.00412	0.004	0.00399	0.00383	0.00383	0.00301	0.00294	0.003	0.00312
49	0.00451	0.00441	0.00434	0.00432	0.00424	0.00418	0.00332	0.00324	0.00334	0.00326
50	0.00494	0.00482	0.00478	0.00475	0.0046	0.00461	0.00455	0.00456	0.00453	0.00447
51	0.00547	0.00532	0.00521	0.00518	0.00517	0.0051	0.00502	0.005	0.00499	0.00494
52	0.00586	0.00585	0.00553	0.0056	0.00563	0.00562	0.00556	0.00555	0.00555	0.00552
53	0.00634	0.00643	0.00616	0.00615	0.00611	0.006	0.00601	0.00602	0.00602	0.00602
54	0.00683	0.00684	0.00667	0.00668	0.00652	0.00657	0.00661	0.00662	0.00662	0.00662
55	0.00748	0.00728	0.00708	0.00722	0.00713	0.0071	0.00713	0.00724	0.00726	0.00716

```

cnn = Sequential()
cnn.add(BatchNormalization())

cnn.add(Conv2D(filters=numberFilters1,
               kernel_size=(filterSize1,filterSize1),
               strides=(1,1),
               padding='valid',
               data_format='channels_last'))
cnn.add(BatchNormalization())
cnn.add(Activation('relu'))

cnn.add(Conv2D(filters=numberFilters2,
               kernel_size=(filterSize2,filterSize2),
               strides=(1,1),
               padding='valid',
               data_format='channels_last'))
cnn.add(BatchNormalization())
cnn.add(Activation('relu'))

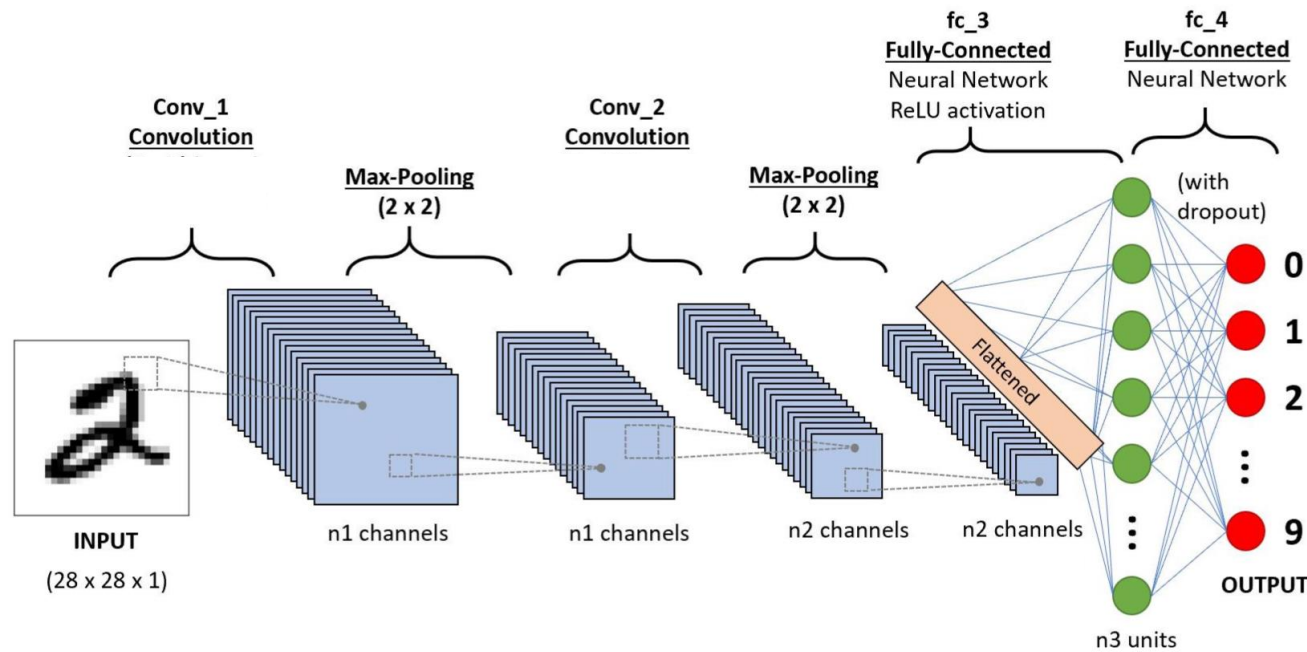
cnn.add(Conv2D(filters=numberFilters3,
               kernel_size=(filterSize3,filterSize3),
               strides=(1,1),
               padding='valid',
               data_format='channels_last'))
cnn.add(BatchNormalization())
cnn.add(Activation('relu'))

cnn.add(Flatten())
cnn.add(Dense(1))
cnn.add(Activation('sigmoid'))
cnn.compile(loss='mean_squared_error', optimizer='sgd')
    
```



See case study 9 at actuarialdatascience.org

Example 2: Detecting anomalies in mortality rates



Layer	input size	output size	$f_1^{(k)}, f_2^{(k)}$
1. Conv.	28×28	$26 \times 26 \times 10$	3, 3
2. Batch norm.	$26 \times 26 \times 10$	$26 \times 26 \times 10$	–
3. ReLU ϕ	$26 \times 26 \times 10$	$26 \times 26 \times 10$	–
4. Max-pooling	$26 \times 26 \times 10$	$13 \times 13 \times 10$	2, 2
5. Conv.	$13 \times 13 \times 10$	$11 \times 11 \times 20$	3, 3
6. Batch norm.	$11 \times 11 \times 20$	$11 \times 11 \times 20$	–
7. ReLU ϕ	$11 \times 11 \times 20$	$11 \times 11 \times 20$	–
8. Max-pooling	$11 \times 11 \times 20$	$10 \times 10 \times 20$	2, 2
9. Conv.	$10 \times 10 \times 20$	$8 \times 8 \times 40$	3, 3
10. Batch norm.	$8 \times 8 \times 40$	$8 \times 8 \times 40$	–
11. ReLU ϕ	$8 \times 8 \times 40$	$8 \times 8 \times 40$	–
12. Max-pooling	$8 \times 8 \times 40$	$4 \times 4 \times 40$	2, 2
13. Flatten	$4 \times 4 \times 40$	640×1	–
14. Fully-conn.	640×1	10×1	–
15. Softmax output	10×1	10×1	–

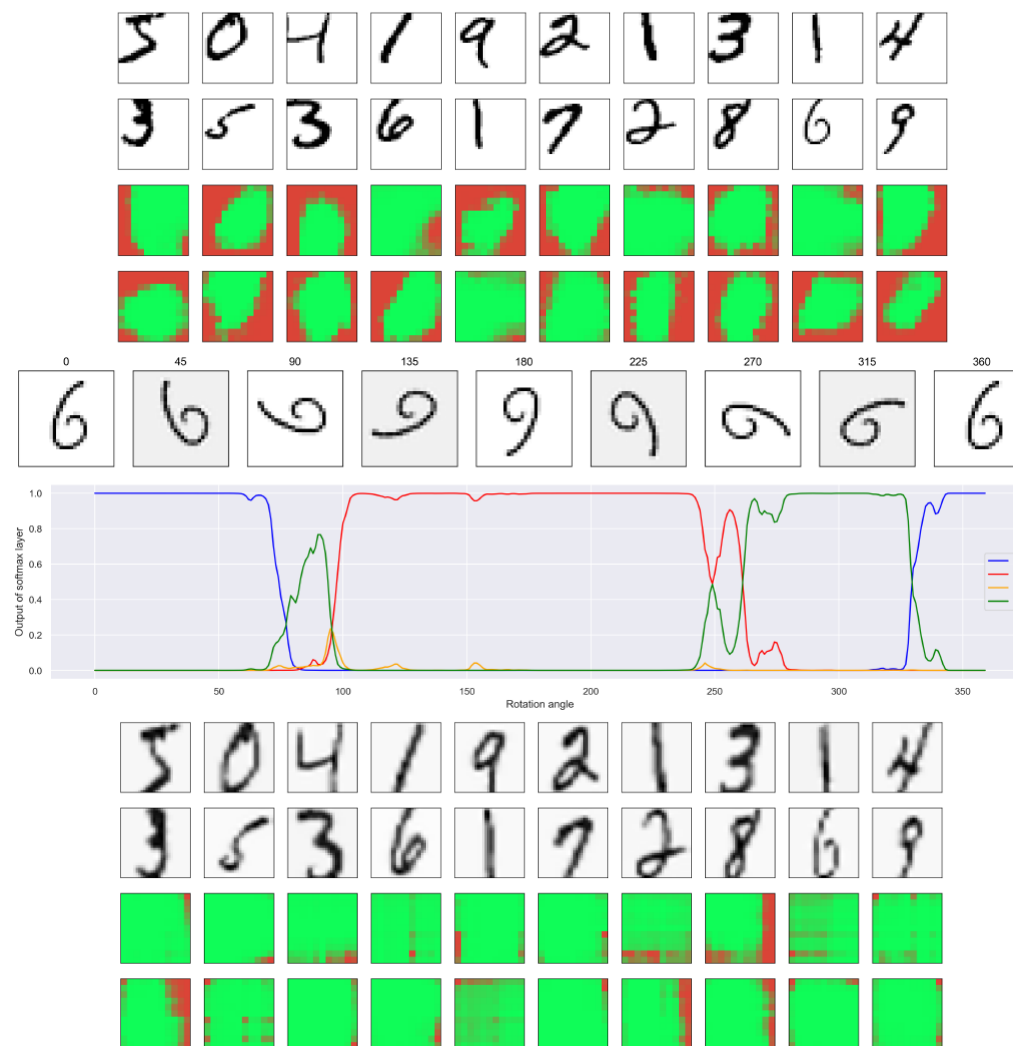
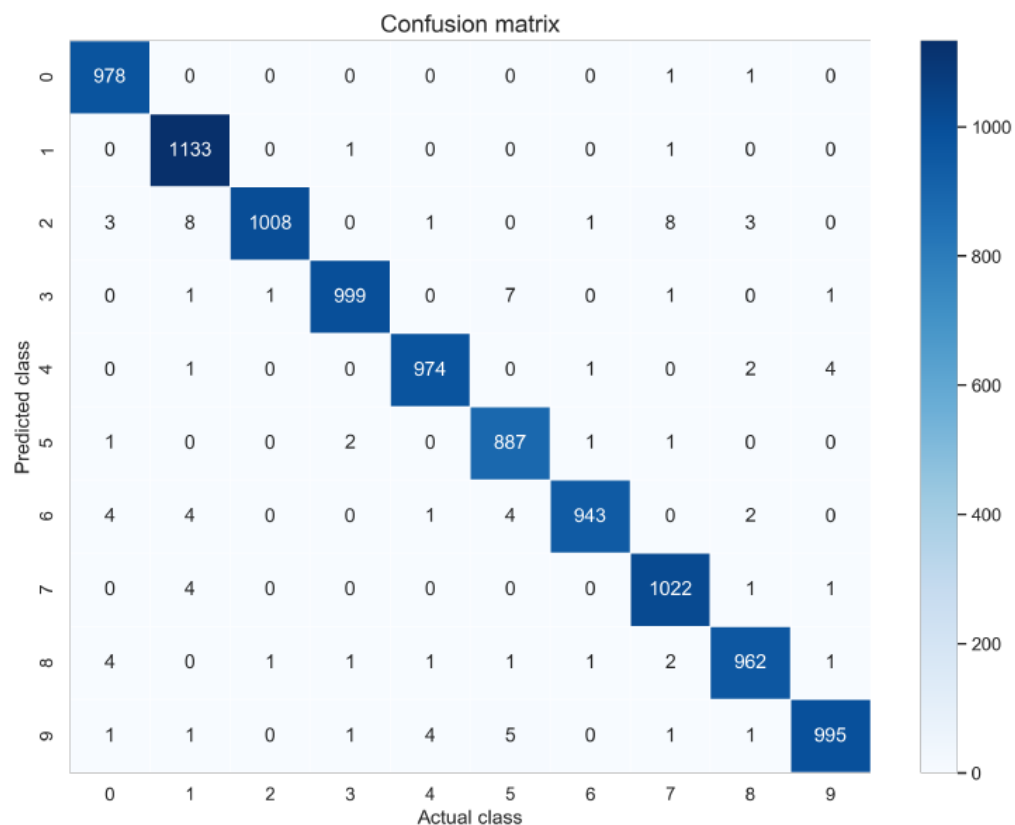
MNIST dataset of 70'000 handwritten digits

The convolutional layer in layer k , applies:

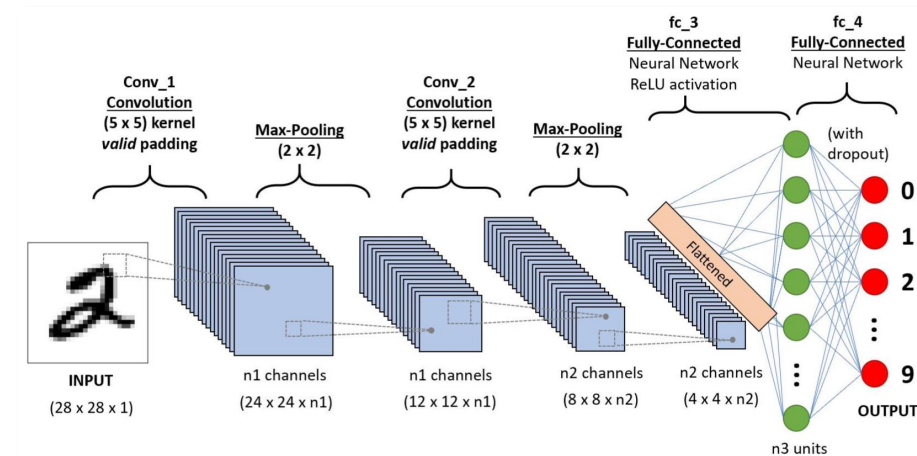
$$\mathbf{x} \mapsto z_{i_1, i_2}^{(k)}(\mathbf{x}) := w_{0,0}^{(k)} + \sum_{j_1=1}^{f_1^{(k)}} \sum_{j_2=1}^{f_2^{(k)}} w_{j_1, j_2}^{(k)} x_{i_1+j_1-1, i_2+j_2-1}$$

15'710 learnable parameters in total

Example 2: Detecting anomalies in mortality rates



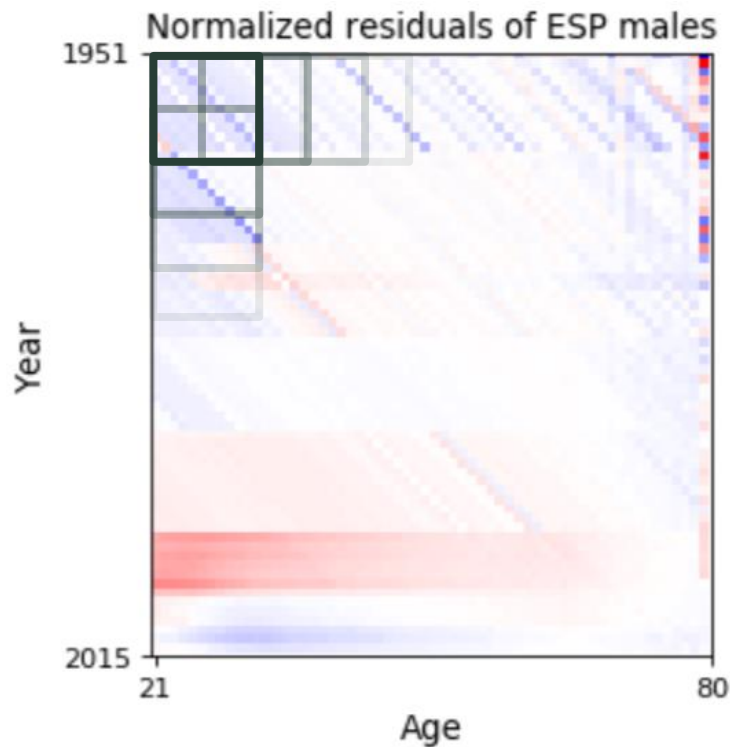
Example 2: Detecting anomalies in mortality rates



Example 2: Detecting anomalies in mortality rates

Model input: 10x10 morality rates (x males/females/difference)

Model output/labels: max abs residual



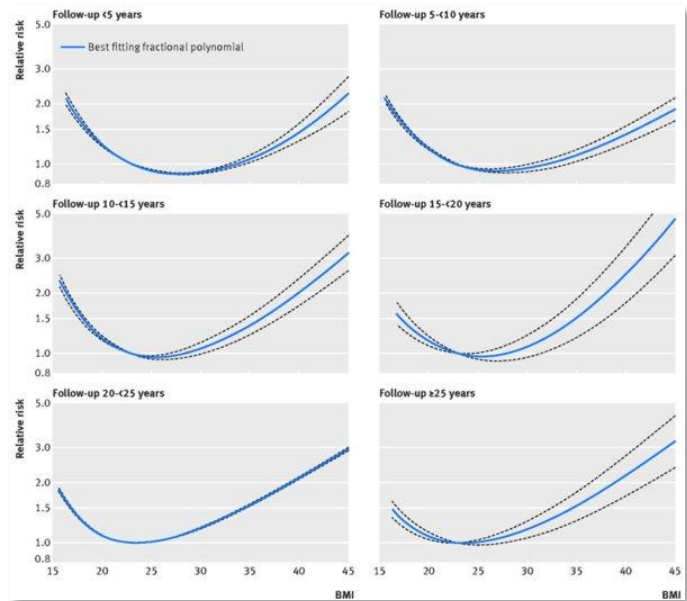
Layer	input size	output size	$f_1^{(k)}, f_2^{(k)}$
1. Batch norm.	$10 \times 10 \times 3$	$10 \times 10 \times 3$	—
2. Conv.	$10 \times 10 \times 3$	$8 \times 8 \times 16$	3, 3
3. Batch norm.	$8 \times 8 \times 16$	$8 \times 8 \times 16$	—
4. ReLU ϕ	$8 \times 8 \times 16$	$8 \times 8 \times 16$	—
5. Conv.	$8 \times 8 \times 16$	$6 \times 6 \times 32$	3, 3
6. Batch norm.	$6 \times 6 \times 32$	$6 \times 6 \times 32$	—
7. ReLU ϕ	$6 \times 6 \times 32$	$6 \times 6 \times 32$	—
8. Conv.	$6 \times 6 \times 32$	$4 \times 4 \times 64$	3, 3
9. Batch norm.	$4 \times 4 \times 64$	$4 \times 4 \times 64$	—
10. ReLU ϕ	$4 \times 4 \times 64$	$4 \times 4 \times 64$	—
11. Flatten	$4 \times 4 \times 64$	$1'024 \times 1$	—
12. Fully-conn.	$1'024 \times 1$	1×1	—
13. Sigmoid output	1×1	1×1	—

24'839 learnable parameters in total

Example 3: Neural networks as an alternative to classical survival models

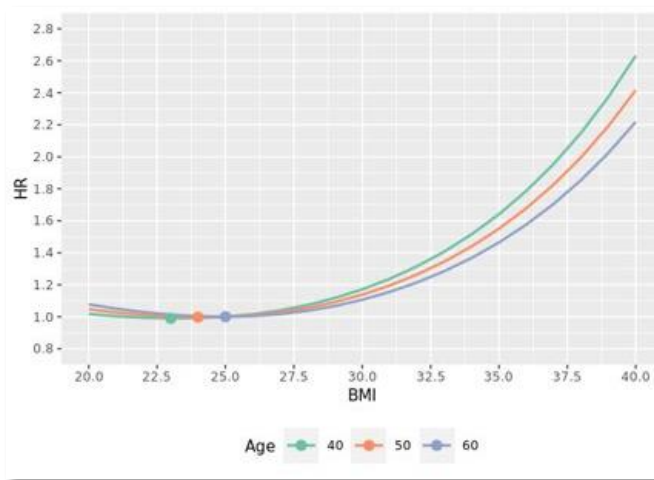
- Develop a **risk score** r that takes as inputs
 - individual **risk factors** R_i like BMI, blood pressure, step counts, etc. (as well as age x , gender g),
 - **insurance product information** p like (e.g. mortality or critical illness, policy duration, regional aspects, etc.), and
 - necessary **expert judgment** D on e.g. long-term trends

and as output first derives *relative risk* with respect to mortality/morbidity.



Source: www.bmj.com/content/353/bmj.i2156

$$r_{x,g,p}: R_1 \times R_2 \times \dots \times R_n \times D \rightarrow (0, \infty)$$



Source: Swiss Re internal project on UK THIN data



Commercial risk scores

Example 3: Neural networks as an alternative to classical survival models

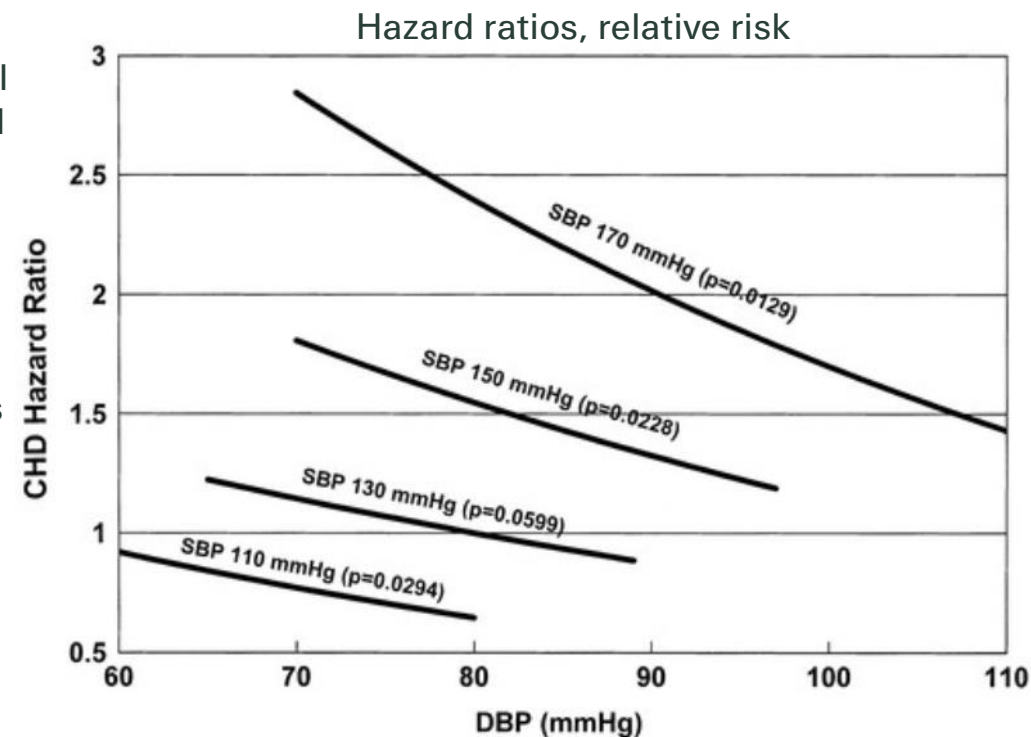
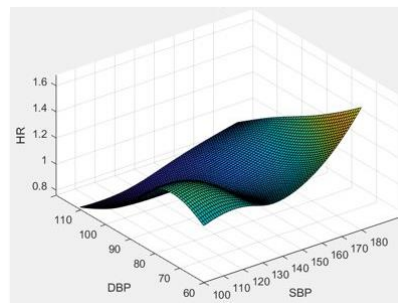
Large longitudinal datasets

ID	Year	BMI	SBP	DBP	...	Death
1	2005	25	130	84	...	0
1	2006	26	136	88	...	0
1	2007	26	126	82	...	0
1	2008	26	128	84	...	0
1	2009	27	130	82	...	1
2	2007	23	116	70	...	0
2	2008	23	122	74	...	0
...	...	...	...	...	...	...

- Cox proportional hazard model
- Accelerated failure time model



- Recurrent neural networks
- Convolutional neural networks



What's in from this for L&H insurance?

Understanding the dependencies and interactions between various mortality/morbidity drivers for underwriting and pricing.

Is Pulse Pressure Useful in Predicting Risk for Coronary Heart Disease?

The Framingham Heart Study

Stanley S. Franklin, Shehzad A. Khan, Nathan D. Wong, Martin G. Larson and Daniel Levy
<https://doi.org/10.1161/01.CIR.100.4.354>, Circulation. 1999;100:354–360

Key take-aways

- Neural networks should be part of the model toolbox of Life & Health actuaries
- There are plenty of model libraries, tutorials, online courses available – and data availability is also improving
- **Try it out...**



3 examples

Thank you!

Any questions?

Contact us



Daniel Meier
L&H R&D Manager
daniel_meier@swissre.com

Follow us





Legal notice

©2022 Swiss Re. All rights reserved. You may use this presentation for private or internal purposes but note that any copyright or other proprietary notices must not be removed. You are not permitted to create any modifications or derivative works of this presentation, or to use it for commercial or other public purposes, without the prior written permission of Swiss Re.

The information and opinions contained in the presentation are provided as at the date of the presentation and may change. Although the information used was taken from reliable sources, Swiss Re does not accept any responsibility for its accuracy or comprehensiveness or its updating. All liability for the accuracy and completeness of the information or for any damage or loss resulting from its use is expressly excluded.