Forecasting Deaths from a Single Cause with Competing Risks

Jim Oeppen\textsuperscript{1}  Søren Kjærgaard\textsuperscript{1}

\textsuperscript{1}Max-Planck Odense Center on the Biodemography of Aging, Odense, Denmark

4th Human Mortality Database Symposium
Berlin, 22nd – 23rd May, 2017
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
One Cause of Death and Competing Risks

Simplex: constrained space

Unconstrained, real space

Geometric Mean Life Table

Forecast Centred dx

Forecast

SVD

Oeppen and Kjærgaard

Forecasting Deaths from a Single Cause with Competing Risks
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations
Deaths by Group of Causes: France, female

Group
- Cardio-Vascular
- Neoplasms
- Other diseases
- Respiratory
- Injury & Poison
- Digestive
- Infectious

Oeppen and Kjærgaard Forecasting Deaths from a Single Cause with Competing Risks
Geometric Mean $d(x)$

<table>
<thead>
<tr>
<th>Cardio-Vascular</th>
<th>Neoplasms</th>
<th>Other diseases</th>
<th>Respiratory</th>
<th>Injury &amp; Poison</th>
<th>Digestive</th>
<th>Infectious</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Oeppen and Kjærgaard Forecasting Deaths from a Single Cause with Competing Risks
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations
SVD: Multiple Causes of Death

The Problem

The Data

CoDa Arithmetic

CoDa SVD

KLd

Model Estimates

Forecast

Correlations

Oeppen and Kjærgaard

Forecasting Deaths from a Single Cause with Competing Risks

<table>
<thead>
<tr>
<th>Cardio–Vascular</th>
<th>Neoplasms</th>
<th>Other diseases</th>
<th>Respiratory</th>
<th>Injury &amp; Poison</th>
<th>Digestive</th>
<th>Infectious</th>
</tr>
</thead>
</table>

Age Effects (scaled)

Rank 1: 82%
Rank 2: 7%
Rank 3: 3%

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks

<table>
<thead>
<tr>
<th>Rank 1: 82%</th>
<th>Rank 2: 7%</th>
<th>Rank 3: 3%</th>
<th>Rank 4: 1% (not used)</th>
</tr>
</thead>
</table>

Age Effects (scaled)

Oeppen and Kjærgaard: Forecasting Deaths from a Single Cause with Competing Risks
SVD: Time Vectors

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations
Kullback–Leibler Divergence

\[ D_{KL}(P \| Q) = \sum_i P_i \log \frac{P_i}{Q_i} \]  

(1)

\[ D_{KL}(P \| Q) \geq 0 \]  

(2)

\[ D_{KL}(n_{dx} \| \hat{n}_{dx}) = \sum_{x=0}^{\omega} n_{dx} \log \frac{n_{dx}}{\hat{n}_{dx}} \]  

(3)
Oeppen and Kjærgaard

Forecasting Deaths from a Single Cause with Competing Risks
### Each Disease versus the Remainder

<table>
<thead>
<tr>
<th>Target &amp; Size</th>
<th>Disease</th>
<th>Rank</th>
<th>Other</th>
<th>Across All $D_{KL}$</th>
<th>Rank</th>
<th>Within Target $D_{KL}$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>0</td>
<td></td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C–V</td>
<td>1</td>
<td>Rem.</td>
<td>0.044</td>
<td>5</td>
<td>0.029</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Neop.</td>
<td>2</td>
<td>Rem.</td>
<td>0.034</td>
<td>4</td>
<td>0.023</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Other D</td>
<td>3</td>
<td>Rem.</td>
<td>0.031</td>
<td>2</td>
<td>0.061</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>I &amp; P</td>
<td>4</td>
<td>Rem.</td>
<td>0.032</td>
<td>3</td>
<td>0.072</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Resp.</td>
<td>5</td>
<td>Rem.</td>
<td>0.063</td>
<td>6</td>
<td>0.130</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Digest.</td>
<td>6</td>
<td>Rem.</td>
<td>0.079</td>
<td>7</td>
<td>0.159</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Inf.</td>
<td>7</td>
<td>Rem.</td>
<td>0.013</td>
<td>1</td>
<td>0.322</td>
<td>7</td>
</tr>
</tbody>
</table>
Neoplasms versus 1 Other and the Remainder

<table>
<thead>
<tr>
<th>Target</th>
<th>Other</th>
<th>Across All</th>
<th>Within Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>Rank</td>
<td>Disease</td>
<td>Rank</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>Rem.</td>
<td>0</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>C–V</td>
<td>1</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>Other D</td>
<td>3</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>I &amp; P</td>
<td>4</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>Resp.</td>
<td>5</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>Digest.</td>
<td>6</td>
</tr>
<tr>
<td>Neop.</td>
<td>2</td>
<td>Inf.</td>
<td>7</td>
</tr>
</tbody>
</table>
Best Model by Complexity Level

<table>
<thead>
<tr>
<th>Target</th>
<th>Competing Risks</th>
<th>$D_{KL}$</th>
<th>Across</th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasms</td>
<td>Remainder</td>
<td></td>
<td>0.034</td>
<td>0.023</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>Other Dis., Remainder</td>
<td></td>
<td>0.047</td>
<td>0.024</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>Other Dis., Inj. &amp; Poison, Remainder</td>
<td></td>
<td>0.050</td>
<td>0.024</td>
</tr>
</tbody>
</table>
Best Model by Complexity Level

<table>
<thead>
<tr>
<th>Target</th>
<th>Competing Risks</th>
<th>Across</th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardio–V. Remainder</td>
<td></td>
<td>0.044</td>
<td>0.029</td>
</tr>
<tr>
<td>Cardio–V. Other Dis., Remainder</td>
<td></td>
<td>0.059</td>
<td>0.029</td>
</tr>
<tr>
<td>Cardio–V. Other Dis., Inj. &amp; Poison, Remainder</td>
<td></td>
<td>0.054</td>
<td>0.028</td>
</tr>
</tbody>
</table>
Outline

1. The Problem
2. The Data
3. Compositional Arithmetic
4. CoDa Lee–Carter and the SVD
5. Kullback–Leibler Divergence
6. Model Estimates
7. Forecast
8. Cause of Death Correlations

Oeppen and Kjærgaard  Forecasting Deaths from a Single Cause with Competing Risks
Deaths by Group of Causes: France, female

Group:
- Cardio-Vascular
- Neoplasms
- Other diseases
- Respiratory
- Injury & Poison
- Digestive
- Infectious

Type:
- Data
- Est
- Fcst

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
Life Expectancy at Birth: female

Data
- Obs

Model
- CoDa Rank 1
- CoDa Rank 3
- Lee−C Rank 1

Decrement
- Single
- Multiple

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
### Centred data: clr transform

<table>
<thead>
<tr>
<th>Category</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardio-Vascular</td>
<td>0.64</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>0.87</td>
</tr>
<tr>
<td>Other diseases</td>
<td>0.64</td>
</tr>
<tr>
<td>Injury &amp; Poison</td>
<td>0.54</td>
</tr>
<tr>
<td>Respiratory</td>
<td>0.61</td>
</tr>
<tr>
<td>Digestive</td>
<td>0.70</td>
</tr>
<tr>
<td>Infectious</td>
<td>0.70</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>0.81</td>
</tr>
<tr>
<td>Other diseases</td>
<td>0.70</td>
</tr>
<tr>
<td>Injury &amp; Poison</td>
<td>0.71</td>
</tr>
<tr>
<td>Respiratory</td>
<td>0.67</td>
</tr>
<tr>
<td>Digestive</td>
<td>0.50</td>
</tr>
<tr>
<td>Infectious</td>
<td>0.81</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>0.59</td>
</tr>
<tr>
<td>Other diseases</td>
<td>0.71</td>
</tr>
<tr>
<td>Injury &amp; Poison</td>
<td>0.67</td>
</tr>
<tr>
<td>Respiratory</td>
<td>0.55</td>
</tr>
<tr>
<td>Digestive</td>
<td>0.76</td>
</tr>
<tr>
<td>Infectious</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Oeppen and Kjærgaard: Forecasting Deaths from a Single Cause with Competing Risks
Fitted Centred data: clr transform

Cardio-Vascular  .69  .86  .79  .63  .89  .83
Neoplasms       .91  .55  .85  .73  .76
Other diseases  .69  .72  .81  .89
Injury & Poison .60  .81  .73
Respiratory     .66  .58
Digestive       .76
Infectious      0   25  50  75  100
Age

Oeppen and Kjærgaard
Forecasting Deaths from a Single Cause with Competing Risks
Compositional residuals: clr transform

<table>
<thead>
<tr>
<th>Cardio-Vascular</th>
<th>.08</th>
<th>.03</th>
<th>.01</th>
<th>.06</th>
<th>.02</th>
<th>.03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasms</td>
<td>.12</td>
<td>-.02</td>
<td>0</td>
<td>.03</td>
<td>-.08</td>
<td></td>
</tr>
<tr>
<td>Other diseases</td>
<td>.05</td>
<td>.10</td>
<td>.04</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury &amp; Poison</td>
<td>-.10</td>
<td>-.01</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>0</td>
<td>-.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digestive</td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>