Additional Evaluations

Results for the Age Group 18 to 52

In Figures 1 and 2 and Tables 1, 2 and 3, we provide some additional results of the clustering-based common age effect model variations for the age group 18 to 52 of the male populations of Austria, Australia, Canada, Switzerland, Denmark, France, the UK, New Zealand, Sweden and the USA.

Generally, some of the algorithms seem to detect stronger dissimilarities in the population-specific age effects compared to the age group 53 to 87, which result in higher numbers of clusters. Also, the clusterings obtained by the algorithms are quite different, reflecting the various approaches and criteria used to measure the dissimilarity between populations. However, there are also some commonalities, for example, (i) Australia, Canada and the USA as well as (ii) Austria, Sweden and the UK exhibit similar age effects according to several clustering methods. Another group that is put together multiple times consists of France and Switzerland, which is not surprising as these countries share a border and have some cultural similarities. Further experiments also demonstrate that all of the algorithms with the slight exception of likelihood-ratio-based clustering exhibit some robustness to small changes in the training data. The CAE model (fitted by Poisson MLE) achieves the best out-of-sample results with respect to bias, MAE and RMSE, while the CAE fuzzy clustering model with $k = 2$ minimizes the MAPE.

We refer to Section 4 of the main text for a more detailed description of the kind of information which can be obtained from the presented figures and tables.

Robustness Check: Out-Of-Sample Results for Different Data

In this section, we provide some tables containing out-of-sample results of our clustering-based CAE models as well as the benchmarks for different data, namely for

- the male populations of the 10 countries investigated by Kleinow (2015) excluding Denmark in Table 4,
- the 21 countries investigated by Li and Lee (2005) in Table 5,
- the male populations of the 10 countries investigated by Kleinow (2015) trained and evaluated on a different time period (up to 2013) than in Section 4 of the main text in Table 6, and
- the male populations of the 10 countries investigated by Kleinow (2015) excluding New Zealand trained and evaluated on a different time period than in Section 4 of the main text, using the most recently available data (up to 2017, downloaded from the Human Mortality Database on February 1, 2021) in Table 7.

Table 1: Comparison of clustering results obtained by different algorithms for males aged 18 to 52 in 10 countries between 1948 and 1987.

Cluster	k-Means	ACF-Based	Likelihood-Ratio-Based (AL)	Fuzzy ML Clustering ($k = 2$)	Fuzzy ML Clustering ($k = 3$)
1	DNK	AUS, AUT, CAN, CHE, FRA, UK, USA	AUS, NZL	AUS, CAN, CHE, NZL, USA	CHE, FRA, NZL
2	NZL	DNK	AUT, DNK	AUT, DNK, FRA, SWE, UK	AUS, CAN, USA
3	CHE, FRA	NZL	CAN	-	AUT, DNK, SWE, UK
4	USA	SWE	FRA	-	-
5	AUS, CAN	-	SWE	-	-
6	AUT, SWE, UK	-	CHE	-	-
7	-	-	UK	-	-
8	-	-	USA	-	-

Table 2: The BIC and its components (maximal log-likelihood $\mathrm{L_{max}}$ and free number of parameters $\mathrm{n_{par}}$) for males aged 18 to 52 in 10 countries between 1948 and 1987. The BIC values for the models fitted by Poisson MLE are not directly comparable to those for the models fitted by SVD/cPCA.

Model	$\mathrm{L_{max}}$	$\mathrm{n_{par}}$	BIC
ACF (SVD)	—	1372	-50477
CAE (cPCA)	—	774	-53834
ILC (SVD)	—	1080	-52638
ILC (MLE)	-69384	1080	149078
CAE (MLE)	-74565	774	156518
CAE(k, C), k-means	-69639	944	148291
CAE(k, C), ACF-based	-74272	876	156907
CAE(k, C), LR (av. linkage)	-69442	1012	148545
CAE Fuzzy, $k = 2$	-70296	816	148382
CAE Fuzzy, $k = 3$ (chosen by BIC)	-69613	856	147397

Table 3: Out-of-sample error measures for males aged 18 to 52 in 10 countries between 1988 and 2007 (trained on 1948 to 1987). Best values in each column are marked in bold.

Model	Bias	MAE	MAPE	RMSE
ACF (SVD)	0.191‰	0.353‰	20.72%	0.508‰
CAE (cPCA)	0.208‰	0.373‰	21.13%	0.545‰
ILC (SVD)	0.231‰	0.391‰	21.99%	0.563‰
ILC (MLE)	0.177‰	0.360‰	21.29%	0.505‰
CAE (MLE)	**0.163**‰	**0.321**‰	20.86%	**0.451**‰
CAE(k, C), k-means	0.175‰	0.360‰	21.42%	0.504‰
CAE(k, C), ACF-based	0.171‰	0.331‰	20.61%	0.476‰
CAE(k, C), LR (av. linkage)	0.170‰	0.354‰	20.82%	0.493‰
CAE Fuzzy, $k = 2$	0.175‰	0.350‰	**20.47%**	0.498‰
CAE Fuzzy, $k = 3$ (chosen by BIC)	0.180‰	0.356‰	20.87%	0.505‰

Table 4: Out-of-sample error measures for males aged 53 to 87 in the 10 countries considered by Kleinow (2015) excluding Denmark between 1988 and 2007 (trained on 1948 to 1987). Best values in each column are marked in bold.

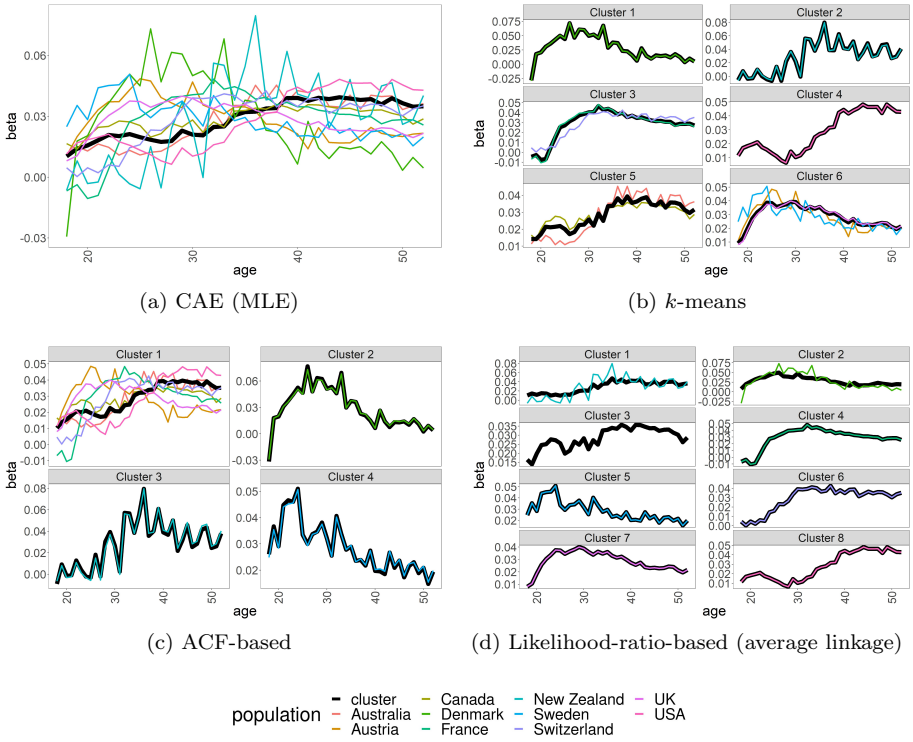
Model	Bias	MAE	MAPE	RMSE
ACF (SVD)	5.788‰	6.684‰	20.59%	9.656‰
CAE (cPCA)	**5.597**‰	6.424‰	19.79%	9.279‰
ILC (SVD)	5.802‰	6.671‰	20.02%	10.031‰
ILC (MLE)	5.946‰	6.712‰	20.11%	10.151‰
CAE (MLE)	5.913‰	6.566‰	19.60%	9.531‰
CAE(k, C), k-means	5.983‰	6.628‰	19.99%	10.028‰
CAE(k, C), ACF-based	5.977‰	6.617‰	19.60%	9.690‰
CAE(k, C), LR (av. linkage)	5.845‰	6.623‰	19.92%	9.784‰
CAE Fuzzy, $k = 2$	5.665‰	**6.271**‰	**19.59%**	**9.114**‰
CAE Fuzzy, $k = 4$ (chosen by BIC)	5.922‰	6.697‰	20.02%	10.198‰

Table 5: Out-of-sample error measures for males aged 53 to 87 in the 21 countries considered by Li and Lee (2005) between 1995 and 2010 (trained on 1959 to 1994). Best values in each column are marked in bold.

Model	Bias	MAE	MAPE	RMSE
ACF (SVD)	**5.294**‰	**6.200**‰	**15.05%**	**9.777**‰
CAE (cPCA)	6.607‰	7.767‰	15.58%	14.176‰
ILC (SVD)	6.368‰	7.049‰	15.43%	12.198‰
ILC (MLE)	6.368‰	6.979‰	15.61%	11.644‰
CAE (MLE)	6.931‰	7.819‰	16.03%	13.809‰
CAE(k, C), k-means	6.446‰	7.025‰	15.60%	11.708‰
CAE(k, C), ACF-based	6.924‰	7.847‰	16.09%	13.896‰
CAE(k, C), LR (av. linkage)	7.161‰	7.771‰	16.31%	13.859‰
CAE Fuzzy, $k = 2$	6.132‰	6.770‰	15.40%	10.905‰
CAE Fuzzy, $k = 5$ (chosen by BIC)	6.362‰	6.955‰	15.53%	11.597‰

Table 6: Out-of-sample error measures for males aged 53 to 87 in the 10 countries considered by Kleinow (2015) between 2000 and 2013 (trained on 1960 to 1999). Best values in each column are marked in bold.

Model	Bias	MAE	MAPE	RMSE
ACF (SVD)	4.617‰	4.919‰	13.32%	7.479‰
CAE (cPCA)	4.603‰	4.919‰	13.02%	7.518‰
ILC (SVD)	**4.435**‰	4.693‰	**12.70%**	7.199‰
ILC (MLE)	4.588‰	4.805‰	12.96%	7.342‰
CAE (MLE)	4.714‰	4.992‰	13.36%	7.531‰
CAE(k, C), k-means	4.541‰	4.754‰	12.86%	7.270‰
CAE(k, C), ACF-based	4.714‰	4.992‰	13.36%	7.531‰
CAE(k, C), LR (av. linkage)	4.640‰	4.833‰	12.94%	7.368‰
CAE Fuzzy, $k = 2$	4.567‰	4.738‰	12.72%	7.258‰
CAE Fuzzy, $k = 3$ (chosen by BIC)	4.501‰	**4.690**‰	12.73%	**7.127**‰



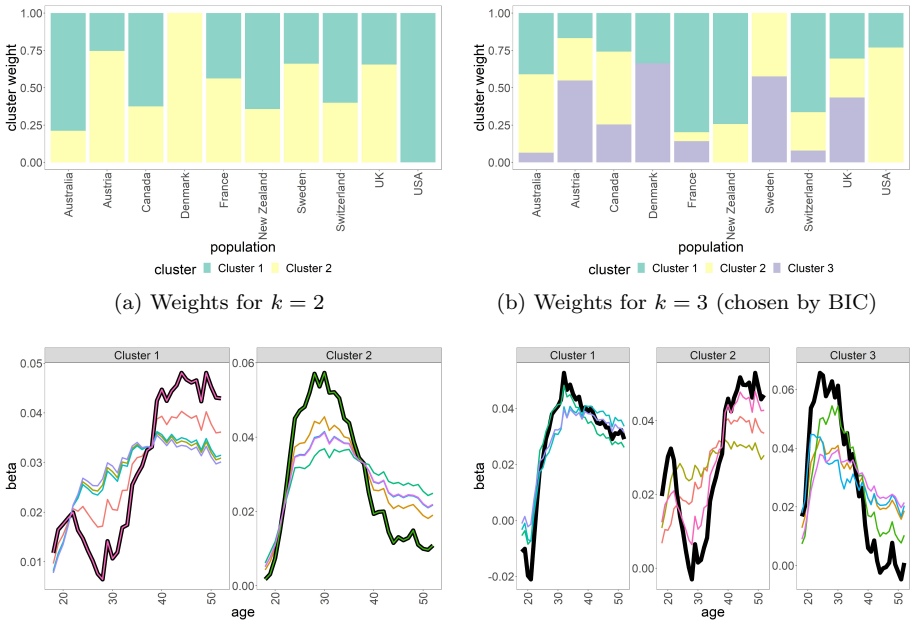
(a) CAE (MLE)

(b) k-means

(c) ACF-based

(d) Likelihood-ratio-based (average linkage)

Figure 1: Age effects by cluster for males aged 18 to 52 in 10 countries between 1948 and 1987 obtained by the CAE model and three of its clustering-based extensions. The age effects of the clusters are displayed in black and the ILC age effects of the populations in different colors.



(a) Weights for $k = 2$

(b) Weights for $k = 3$ (chosen by BIC)

(c) Age effects for $k = 2$

(d) Age effects for $k = 3$ (chosen by BIC)

Figure 2: Weights $\omega^{i,l}$ (top) and cluster-specific age effects $\left(\beta_x^l\right)_x$ as well as fitted population-specific age effects $\left(\sum_{l=1}^k \omega^{i,l}\beta_x^l\right)_x$ (bottom) of the fuzzy maximum likelihood clustering algorithm for males aged 18 to 52 in 10 countries between 1948 and 1987.

Table 7: Out-of-sample error measures for males aged 53 to 87 in the 10 countries considered by Kleinow (2015) excluding New Zealand between 1998 and 2017 (trained on 1958 to 1997). Best values in each column are marked in bold.

Model	Bias	MAE	MAPE	RMSE
ACF (SVD)	5.571‰	5.881‰	16.34%	8.967‰
CAE (cPCA)	5.746‰	6.087‰	16.57%	9.318‰
ILC (SVD)	**5.418**‰	5.744‰	**16.30%**	8.704‰
ILC (MLE)	5.576‰	5.864‰	16.67%	8.881‰
CAE (MLE)	5.798‰	6.105‰	16.91%	9.228‰
CAE(k, C), k-means	5.441‰	**5.724**‰	16.54%	**8.593**‰
CAE(k, C), ACF-based	5.689‰	6.052‰	16.91%	9.071‰
CAE(k, C), LR (av. linkage)	5.585‰	5.840‰	16.78%	8.883‰
CAE Fuzzy, $k = 2$	5.594‰	5.815‰	16.46%	8.829‰
CAE Fuzzy, $k = 3$ (chosen by BIC)	5.513‰	5.751‰	16.46%	8.662‰