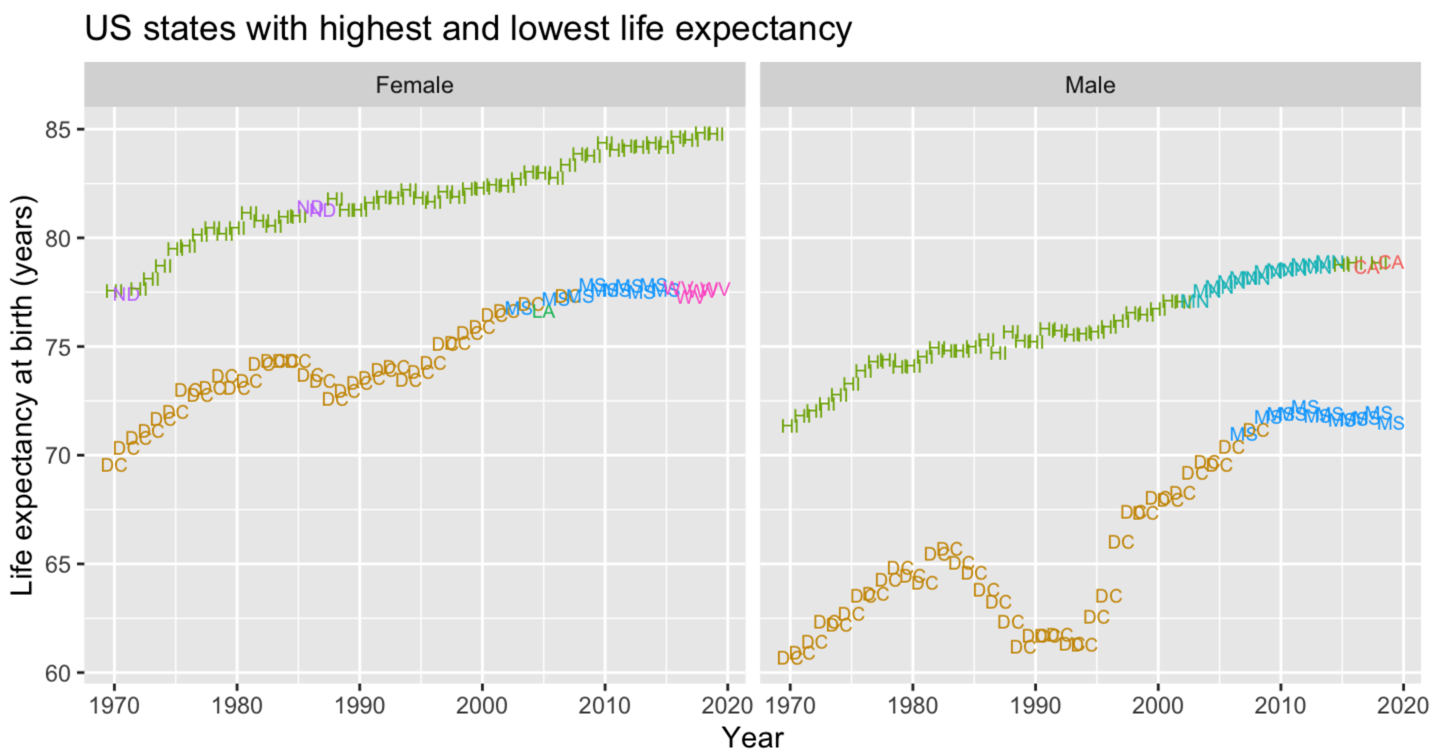
Jointly Estimating Subnational Mortality for Multiple Populations

Monica Alexander, Statistical Sciences and Sociology, University of Toronto. (Joint with Ameer Dharamshi, Celeste Winnat, and Magali Barbieri)

Stellar Stats Workshop, May 26, 2022

Motivation

- Interested in geographic disparities in mortality and also differences across key demographic groups (sex, race/ethnicity)
- Subpopulations in a particular area are likely to have shared experience of mortality risk



Motivation

Motivating example is mortality in US counties. Problems:

- ~3500 counties, some <1000 people. Subgroups have small population counts
- Large stochastic variation, observed death rates are erratic (sometimes zero); underlying mortality risk is often unclear

Goal: build a statistical model that

- takes advantage of strong regularities in mortality across age
- shares information about mortality patterns across space and time
- accounts for subgroup correlations in mortality

Data

This project is motivated by our work extending the Human Mortality Database (HMD) to US counties. Accordingly, we rely on the following US data:

- County-level death and population counts disaggregated by demographic characteristics (in this example: age, sex)
- State-level death and population counts by age and sex
- Data interval: 1982-2018

Methods

We use a Bayesian hierarchical principal components model:

$$y_{a,g,c,t} \sim \text{Poisson}\left(P_{a,g,c,t} \cdot m_{a,g,c,t}\right)$$

$$\log\left(m_{a,g,c,t}\right) = \sum_{i=1}^{K} \left(\beta_{i,g,c,t} \cdot Y_{i,a}\right) + u_{a,g,c,t}$$

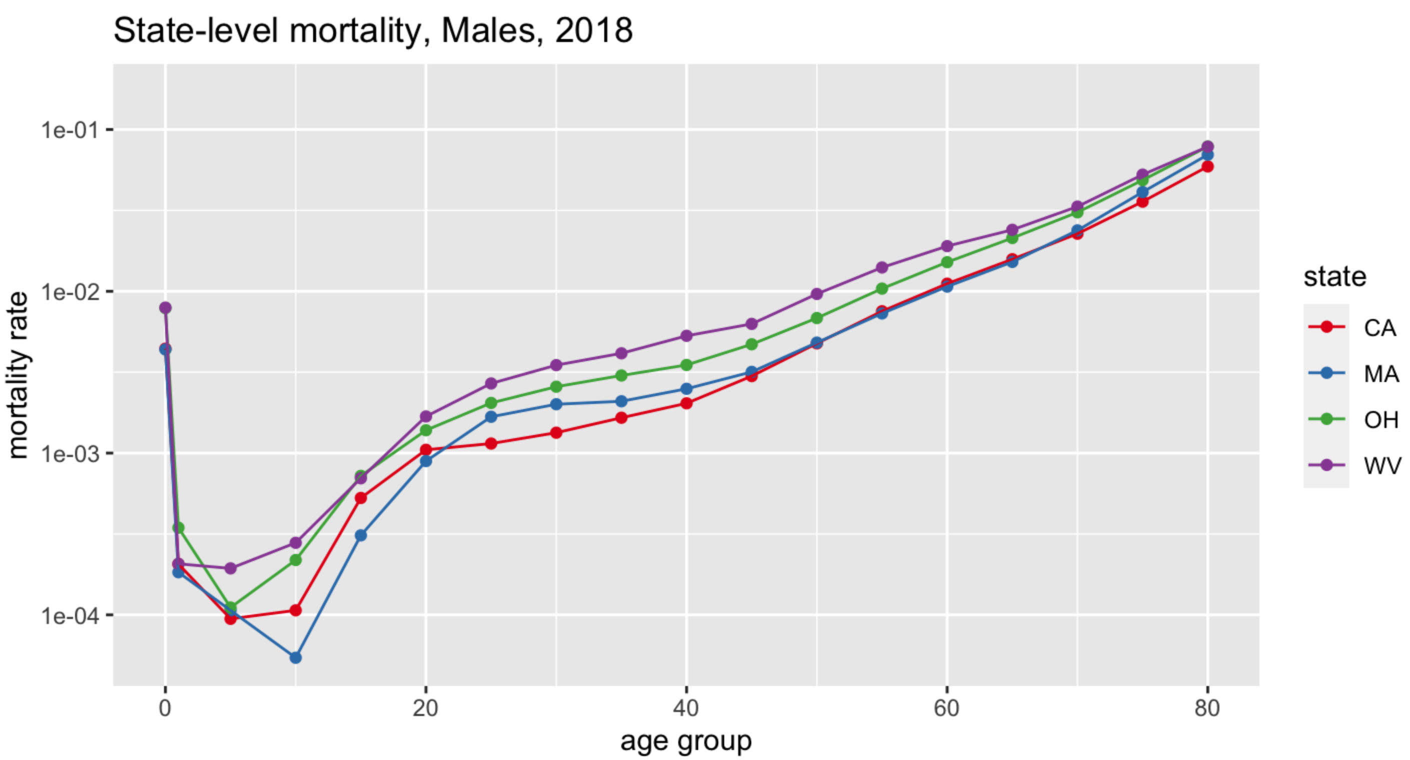
where:

- y is deaths, P is population, m is the mortality rate,
- $Y_{i,a}$ are principal components, β are the principal component coefficients, and u allows for overdispersion,

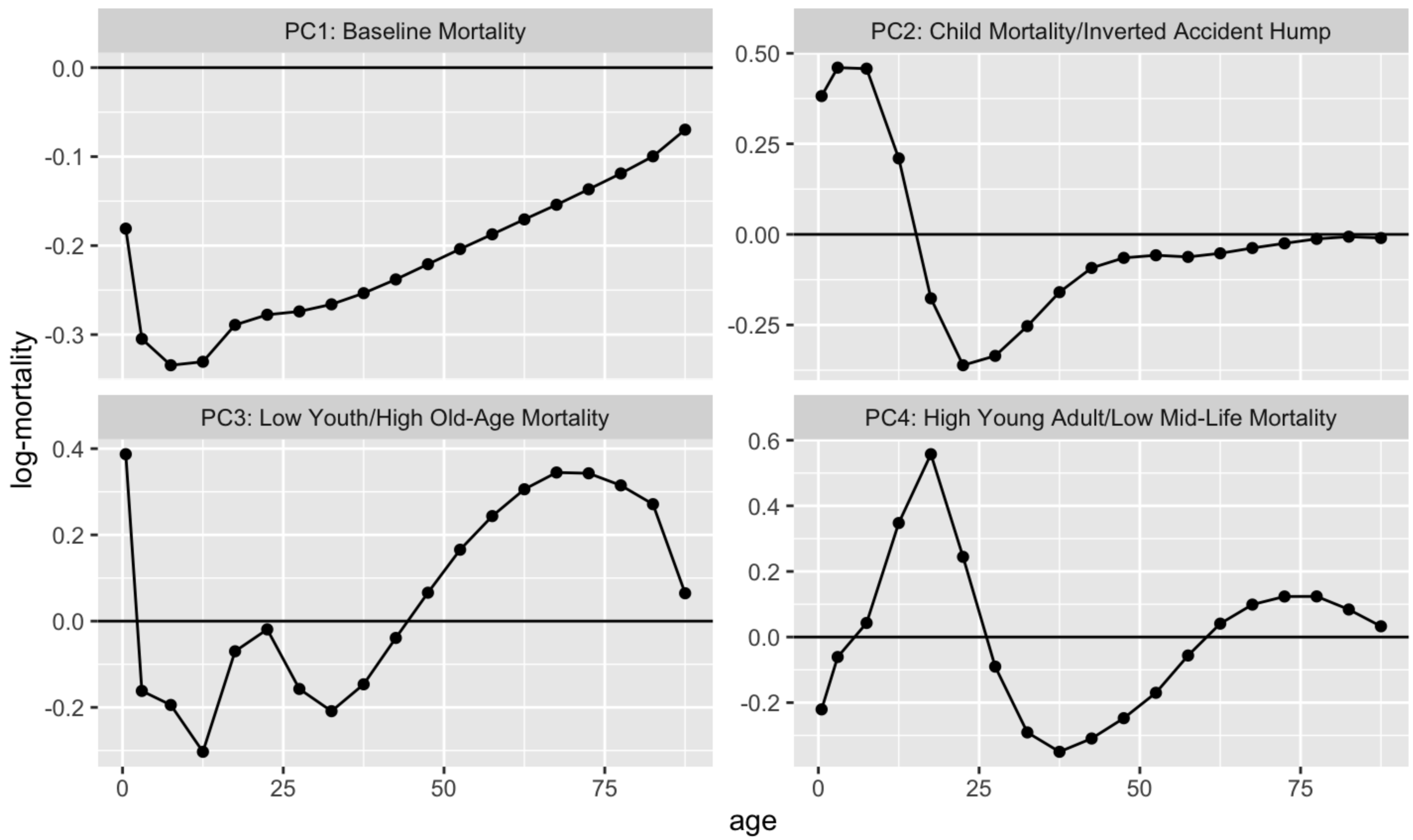
for age group a, subpopulation g, county c, year t, and principal component i

Regularities over age

- Human mortality has characteristic 'J' shape
- Principal components $Y_{i,a}$ are derived from performing SVD on mortality rates at state level
- These explain the main dimensions of variation in mortality across age
- The β coefficients shift up or down based on county-specific observations



Principal Components



Estimating PC Coefficients (β)

PC coefficients are decomposed into state (μ) and county (ω) pieces:

$$\log\left(\mu_{a,g,c,t}\right) = \sum_{i=1}^{K} \left(\beta_{i,g,c,t} \cdot Y_{i,a}\right) + u_{a,g,c,t}$$

$$\begin{pmatrix} \beta_{i,1,c,t} \\ \dots \\ \beta_{i,G,c,t} \end{pmatrix} = \begin{pmatrix} \mu_{\beta_{i,1,t}} \\ \dots \\ \mu_{\beta_{i,G,t}} \end{pmatrix} + \begin{pmatrix} \omega_{i,1,c,t} \\ \dots \\ \omega_{i,G,c,t} \end{pmatrix}, \quad i = 1, \dots, K$$

$$\begin{pmatrix} \omega_{i,1,c,t} \\ \dots \\ \omega_{i,G,c,t} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_G, \sigma_{\beta_{i,t}} \mathbf{1}_G L_{i,t}^{(\beta)} L_{i,t}^{(\beta)T} \mathbf{1}_G \sigma_{\beta_{i,t}}\right)$$

A similar specification is used for the u term.

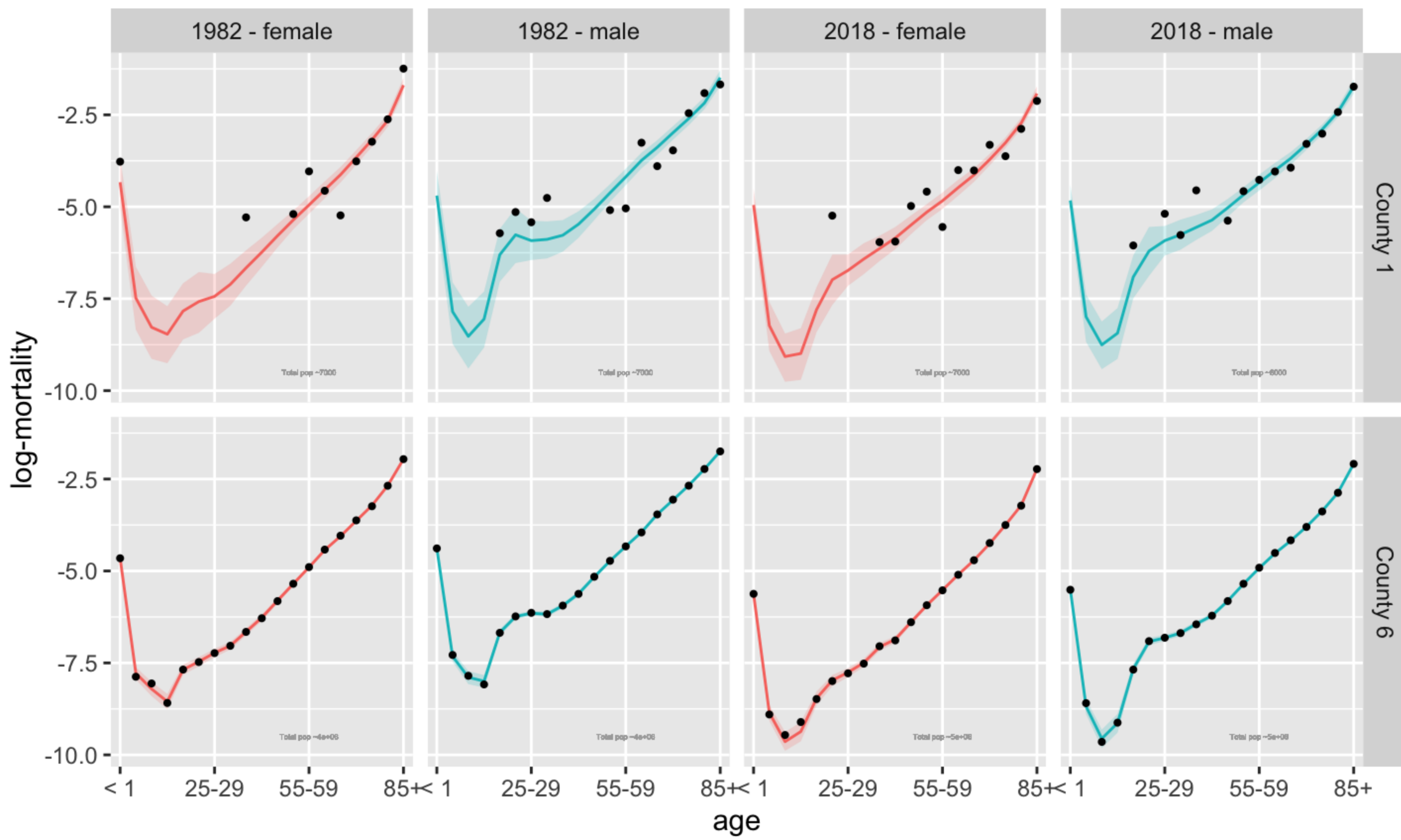
Smoothing over time

- Underlying mortality tends to change gradually over time
- State means are smoothed over time by penalizing the second differences

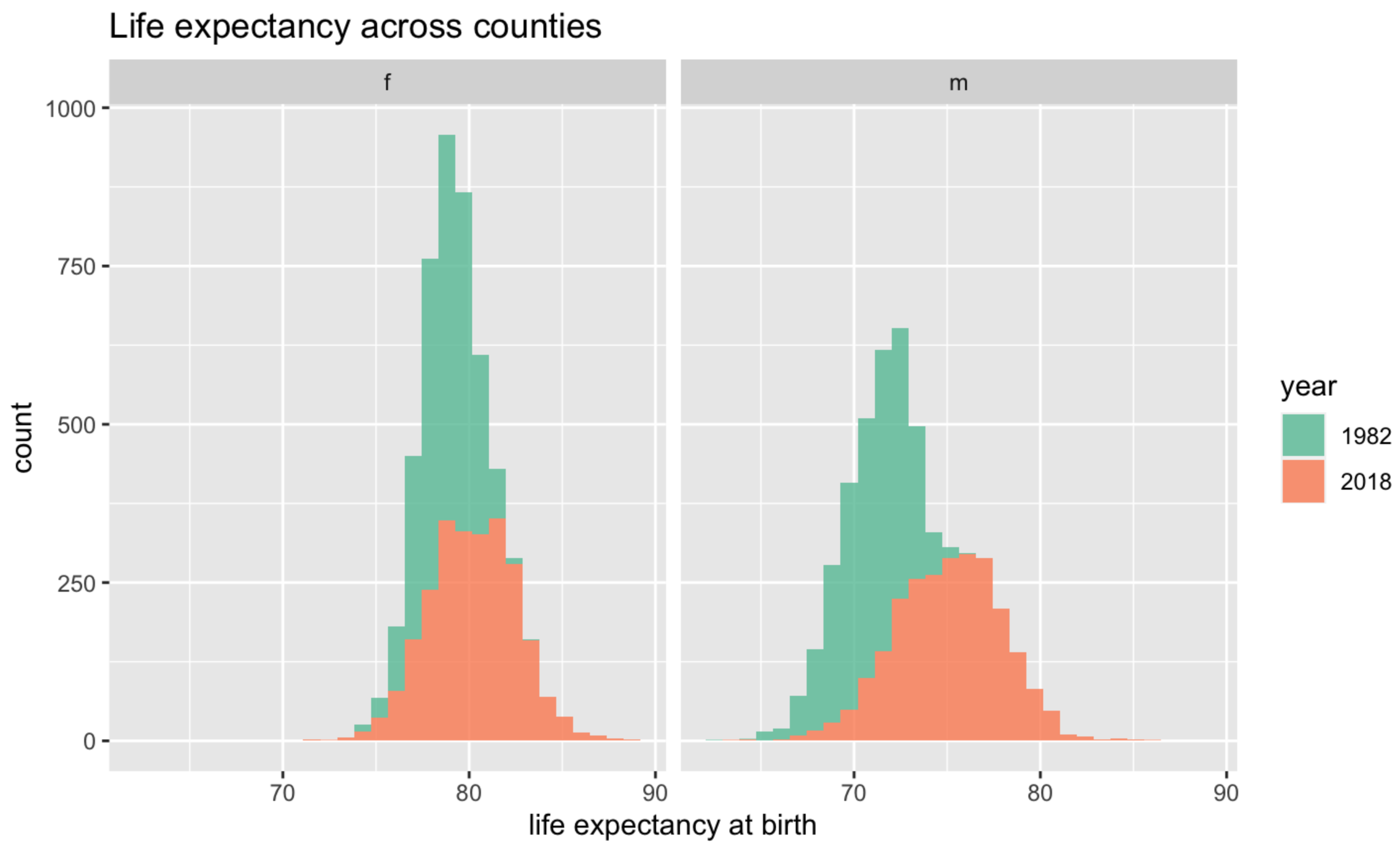
$$\mu_{\beta_{i,s,t}} \sim \mathcal{N}\left(2 \cdot \mu_{\beta_{i,s,t-1}} - \mu_{\beta_{i,s,t-2}}, \sigma_{\mu_{\beta_i}}\right)$$

- Smoothing at the state level allows flexibility for counties to experience irregular mortality patterns driven by localized events such as a natural disaster.

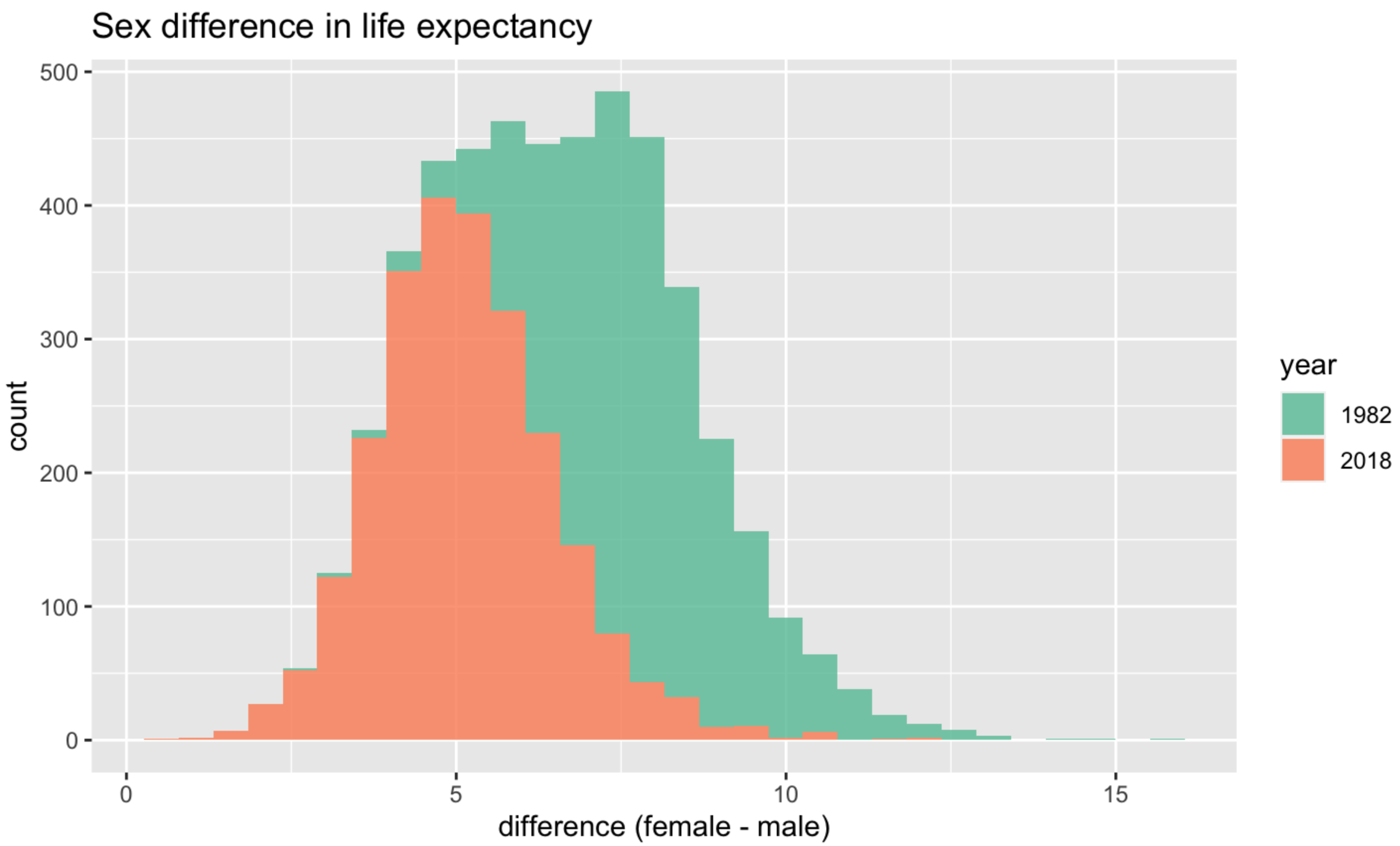
Illustrative Results



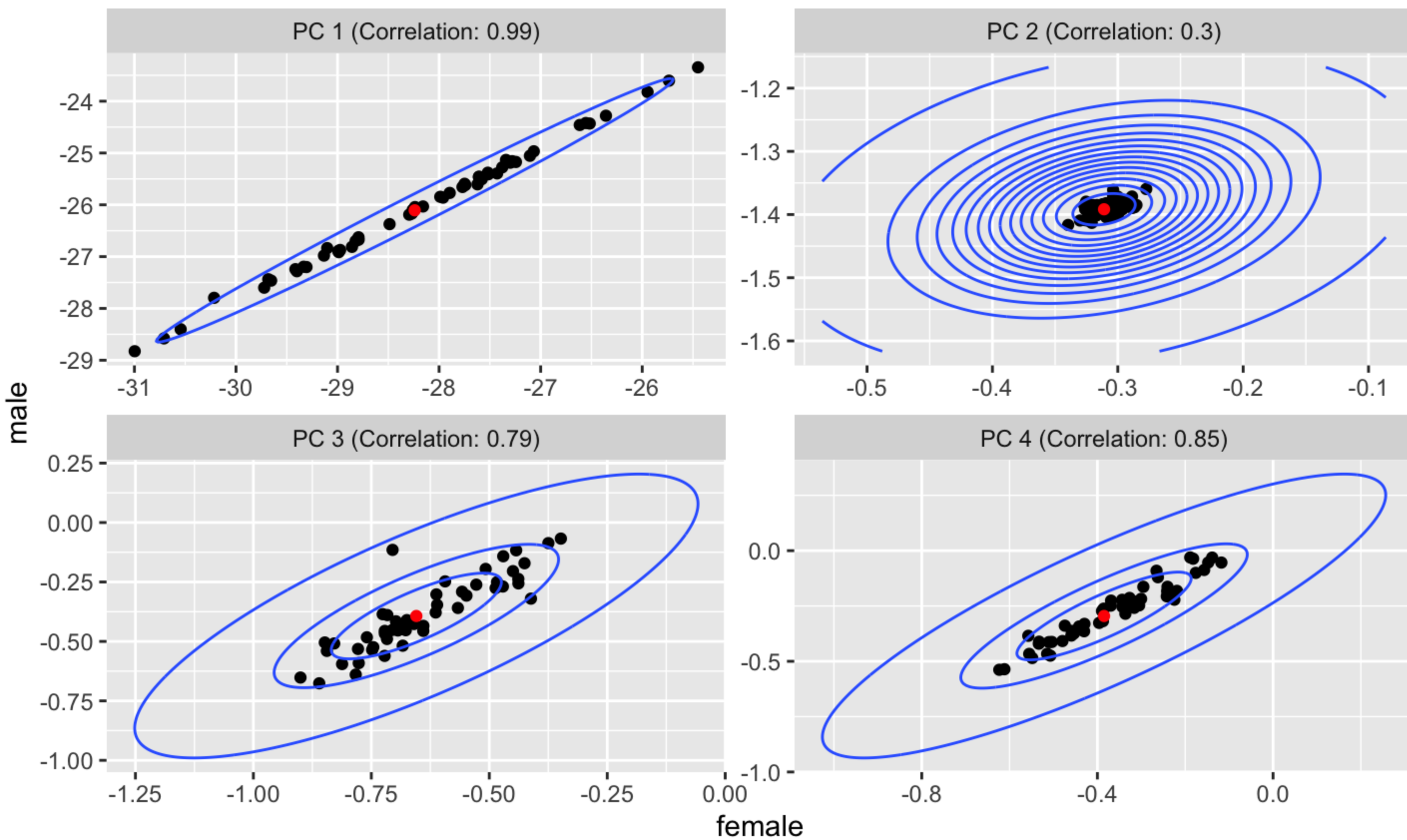
Changing distributions across space



Changes in the distribution of sex disparities



Coefficient Dispersion



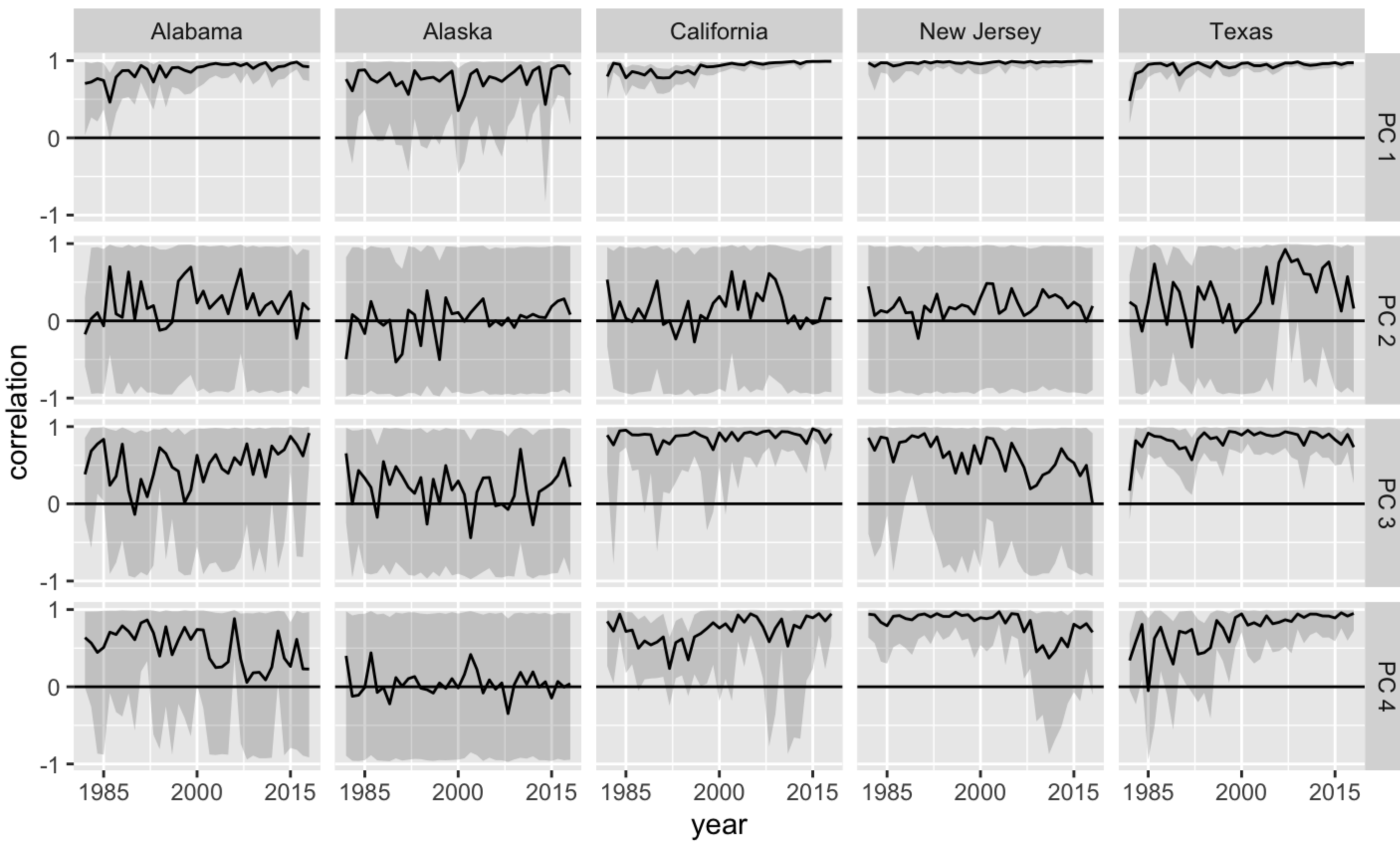
Summary

- Flexible framework that allows mortality risk to be estimated at the subnational area jointly for key subpopulations
- First to explicitly account for shared experience of mortality patterns across groups
- Time series structure allows for future projections
  - Useful for e.g. excess death calculations
- In addition to producing estimates (and uncertainty) for mortality rates; auxiliary parameters give insights into mortality disparities:
  - e.g. Alaska has insignificant correlations suggesting limited interactions between male/female mortality patterns
  - Changes over time have implications for the study of convergence/divergence of mortality by subgroup.

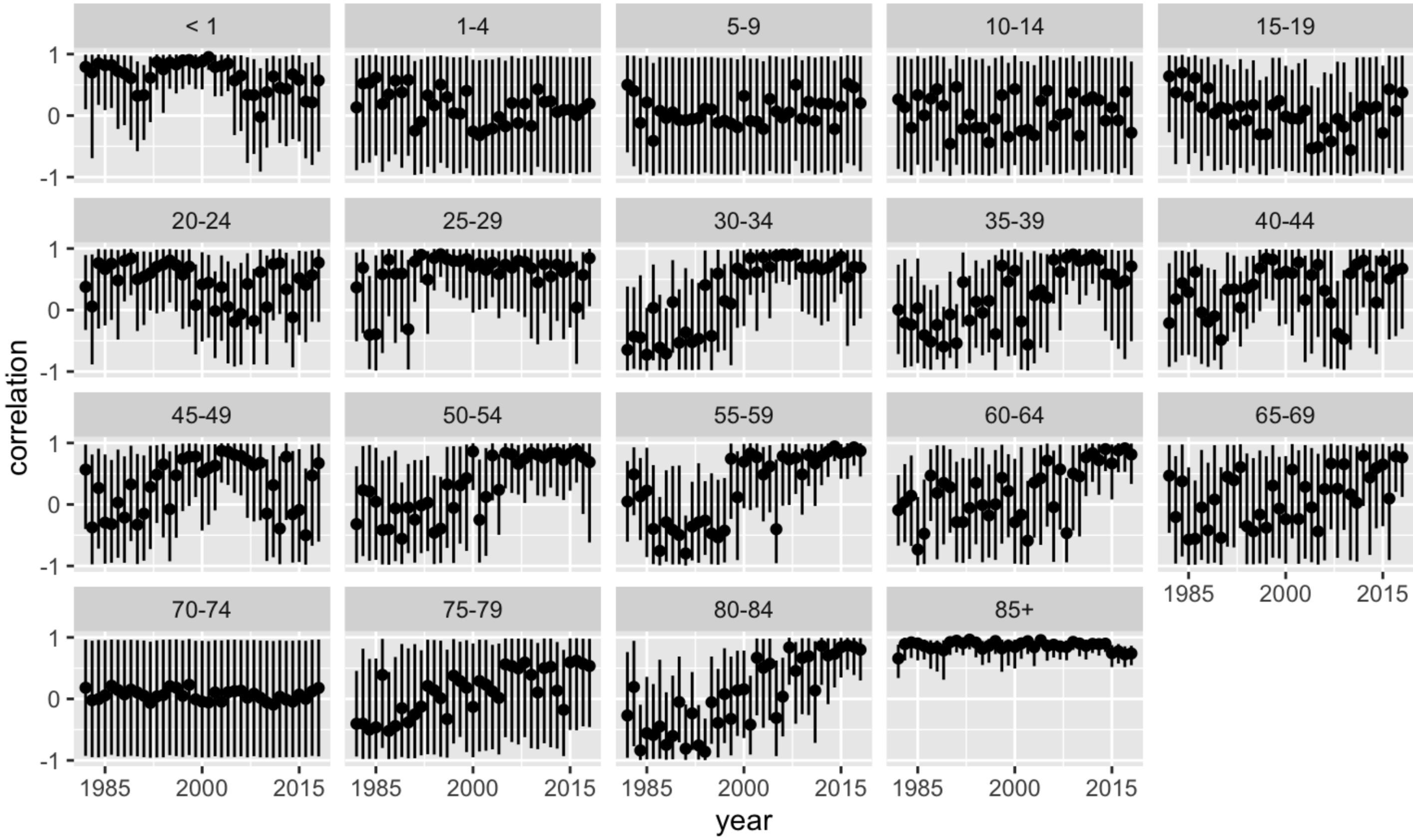
Thanks!

Appendix

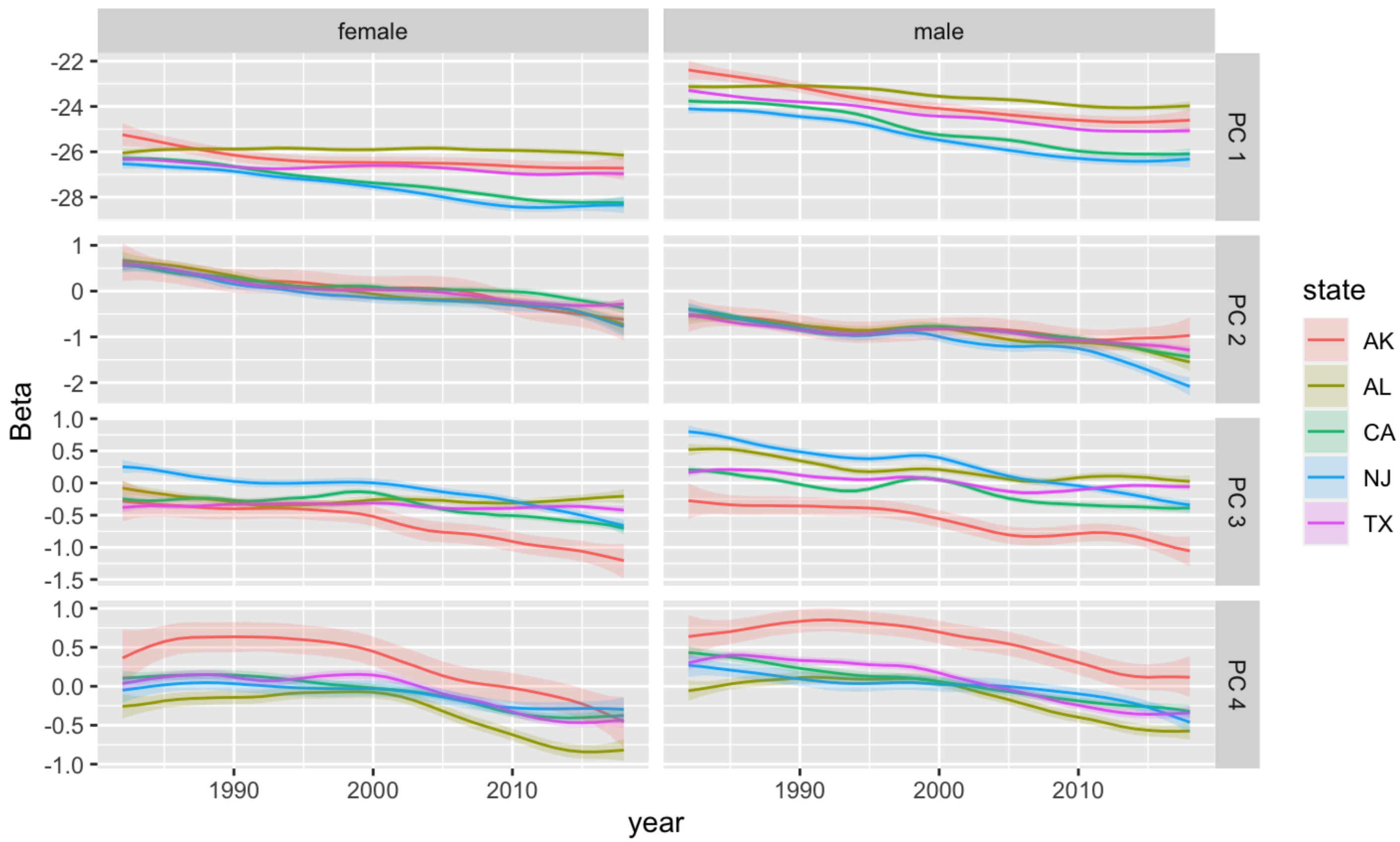
Correlations Over Time



Overdispersion Correlations



State Means


female
male
PC 1
PC 2
PC 3
PC 4
-22
-24
-26
-28
1
0
-1
-2
1.0
0.5
0.0
-0.5
-1.0
-1.5
1.0
0.5
0.0
-0.5
-1.0
1990
2000
2010
1990
2000
2010
year
Beta
state
AK
AL
CA
NJ
TX

State Gaps


PC 1
PC 2
PC 3
PC 4
Mu Beta Gap (female - male)
-2.0
-2.5
-3.0
1.5
1.0
0.5
0.0
0.4
0.0
-0.4
-0.8
0.0
-0.5
-1.0
1990
2000
2010
1990
2000
2010
year
state
AK
AL
CA
NJ
TX

Out-of-Sample Results

State	Model	Cov80	Cov90	Cov95	MAD	MSE
Alaska	indep	0.884	0.943	0.973	1.931	14.447
Alaska	joint	0.887	0.946	0.974	1.899	12.614
California	indep	0.843	0.920	0.961	9.747	2004.650
California	joint	0.846	0.923	0.963	8.568	1194.914
Louisiana	indep	0.873	0.939	0.970	3.054	40.795
Louisiana	joint	0.874	0.940	0.972	2.921	30.257
New Jersey	indep	0.843	0.925	0.963	8.388	298.082
New Jersey	joint	0.850	0.930	0.964	7.838	238.458
Texas	indep	0.878	0.940	0.970	3.232	92.449
Texas	joint	0.879	0.941	0.970	3.030	61.618

In-Sample Results

State	Model	Cov80	Cov90	Cov95	MAD	MSE
Alaska	indep	0.934	0.976	0.991	1.449	4.992
Alaska	joint	0.934	0.976	0.991	1.456	5.024
California	indep	0.932	0.975	0.990	3.695	36.909
California	joint	0.929	0.973	0.989	3.811	40.928
Louisiana	indep	0.923	0.970	0.988	2.195	12.224
Louisiana	joint	0.920	0.968	0.987	2.221	12.597
New Jersey	indep	0.954	0.988	0.996	4.112	41.623
New Jersey	joint	0.947	0.984	0.995	4.257	45.116
Texas	indep	0.929	0.971	0.988	2.010	11.365
Texas	joint	0.925	0.970	0.987	2.052	12.077