# MULTI-POPULATION MORTALITY MODELLING

## Danish data and Basis Risk

Andrew Cairns

Heriot-Watt University, Scotland

and

The Maxwell Institute, Edinburgh

Joint work with

David Blake, Kevin Dowd, Malene Kallestrup-Lamb and Carsten Rosenskjold

## Plan

- Motivation and challenges

- Danish males data

  10 sub-populations grouped by affluence

- An extended CBD multi-population model

- Bayesian implementation and results

# Motivation for multi-population modelling

## A: Risk assessment

- Multi-country (e.g. consistent demographic projections)
- Males/Females (e.g. consistent demographic projections)
- Socio-economic subgroups (e.g. blue or white collar)
- Smokers/Non-smokers
- Annuities/Life insurance
- Limited data $\Rightarrow$ learn from other populations

## Motivation for multi-population modelling

## B: Risk management for pension plans and insurers

- Retain systematic mortality risk; versus:
- 'Over-the-counter' deals (e.g. longevity swap)
- Standardised mortality-linked securities
  - linked to national mortality index
  - $< 100\%$ risk reduction: basis risk

## Challenges

- Data availability
- Data quality and depth
- Model complexity
  - single population models can be complex
  - 2-population versions are more complex
  - multi-pop ......
- Multi-population modelling requires
  - (fairly) simple single-population models
  - simple dependencies between populations

## A New Case Study and a New Model

- Sub-populations differ from national population
  - socio-economic factors
  - geographical variation
  - other factors
- Denmark
  - High quality data on ALL residents
  - 1981-2005 available
  - Can subdivide population using covariates on the database

## Danish Data

- *What can we learn from Danish data that will inform us about other populations?*

- Key covariates
  - Wealth
  - Income
- *Affluence* $=$ Wealth$+15\times$Income

## Problem

- High income $\Rightarrow$ "affluent" *and healthy* BUT
- Low income $\not\Rightarrow$ not affluent, poor health
- High wealth $\Rightarrow$ "affluent" *and healthy* BUT
- Low wealth $\not\Rightarrow$ not affluent, poor health

Solution: use a combination

- Affluence, $A =$ wealth $+ K \times$ income
- $K = 15$ seems to work well *statistically* as a predictor
- Low affluence, $A$, predicts poor mortality

## Subdividing Data

- Males resident in Denmark for the previous 12 months
- Divide population in year $t$
  - into 10 equal sized Groups (approx)
  - using *affluence*, $A$
- Individuals can change groups up to age 67
- Group allocations are locked down at age 67

(better than not locking down at age 67)

## Subdivided Data

- Exposures $E^{(i)}(t,x)$ for groups $i = 1, \ldots, 10$

  range from over 4000 down to 20

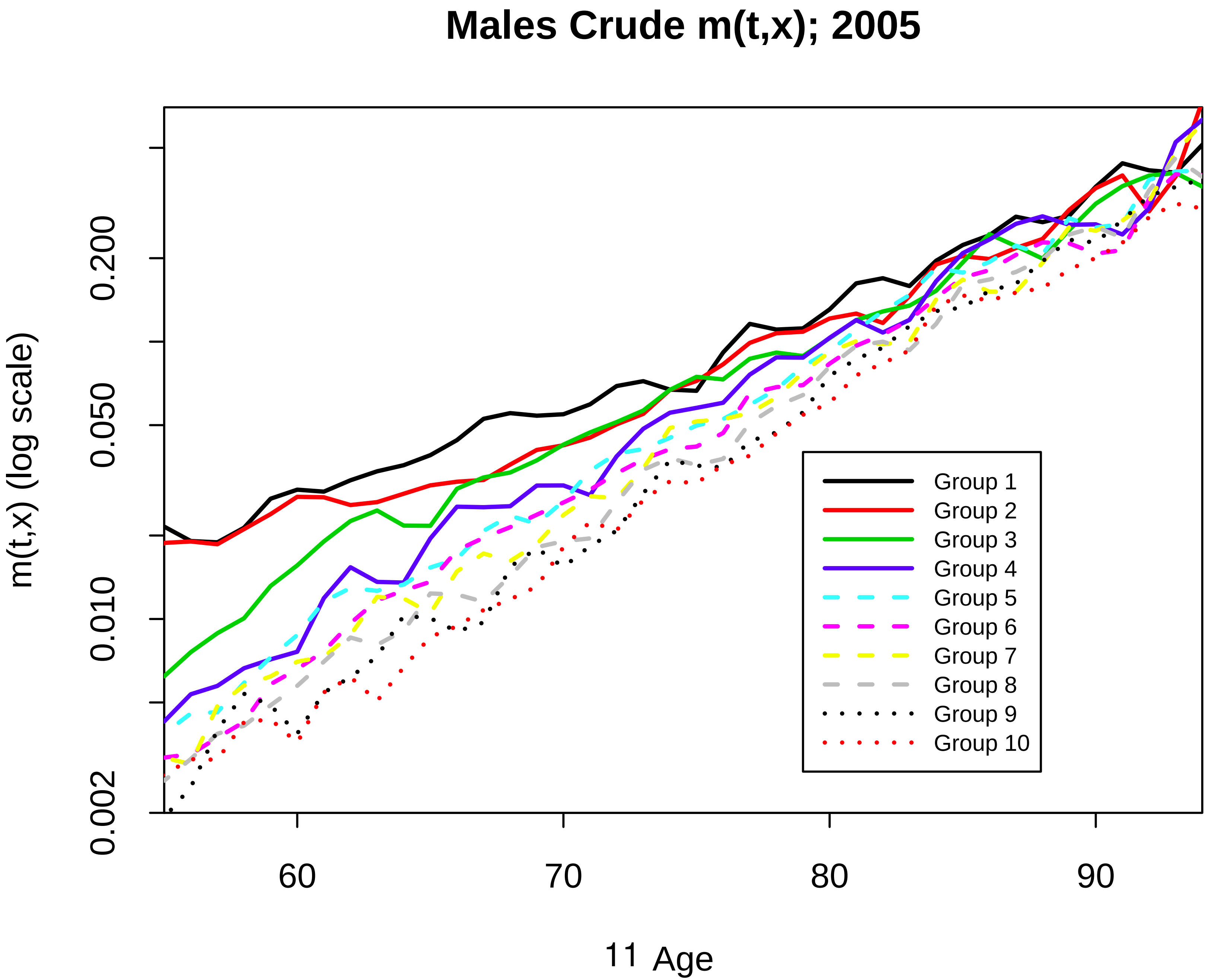
- Deaths $D^{(i)}(t,x)$

  range from 150 down to 6

- Crude death rates $\hat{m}^{(i)}(t,x) = D^{(i)}(t,x)/E^{(i)}(t,x)$

- Small groups $\Rightarrow$ Poisson risk is important

# Crude death rates 2005



## Modelling the death rates, $m_k(t,x)$

$m^{(k)}(t,x) =$ pop. $k$ death rate in year $t$ at age $x$

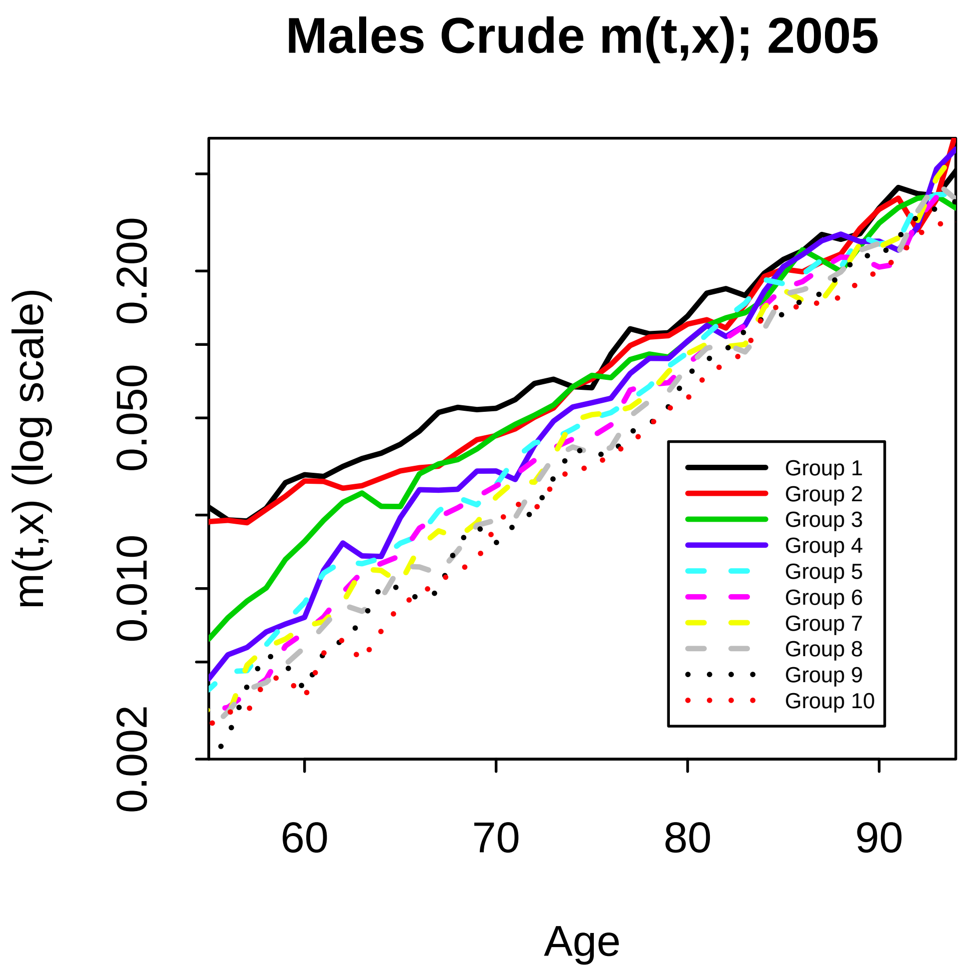
Population $k$, year $t$, age $x$

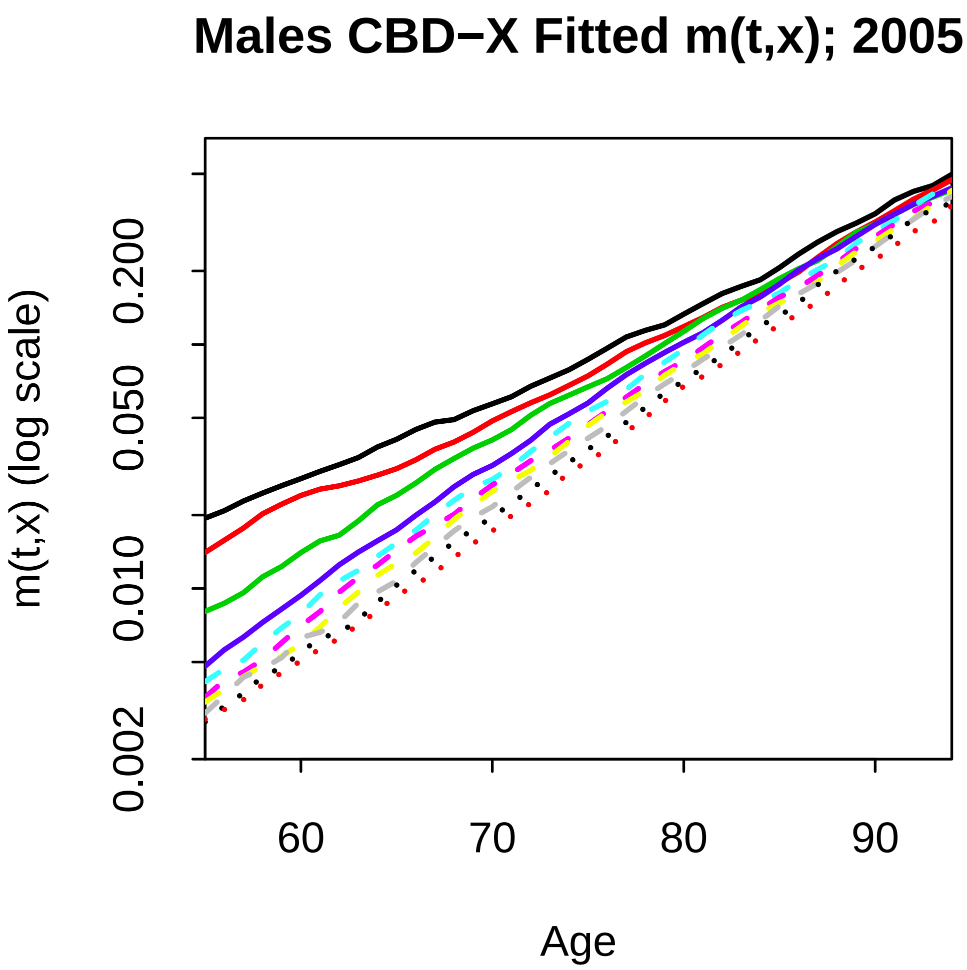
$$\log m^{(k)}(t,x) = \beta^{(k)}(x) + \kappa_1^{(k)}(t) + \kappa_2^{(k)}(t)(x-\bar{x})$$

(Extended CBD with a non-parametric base table, $\beta^{(k)}(x)$)

- 10 groups, $k = 1,\ldots,10$ (low to high affluence)
- 21 years, $t = 1985,\ldots,2005$
- 40 ages, $x = 55,\ldots,94$

# Model-Inferred Underlying Death Rates 2005



## Modelling the death rates, $m_k(t, x)$

$$\log m^{(k)}(t, x) = \beta^{(k)}(x) + \kappa_1^{(k)}(t) + \kappa_2^{(k)}(t)(x - \bar{x})$$

- Model fits the 10 groups well without a cohort effect
- Non-parametric $\beta^{(k)}(x)$ is essential to preserve group rankings
  - Rankings are evident in crude data
  - *"Bio-demographical reasonableness"*:

more affluent $\Rightarrow$ healthier

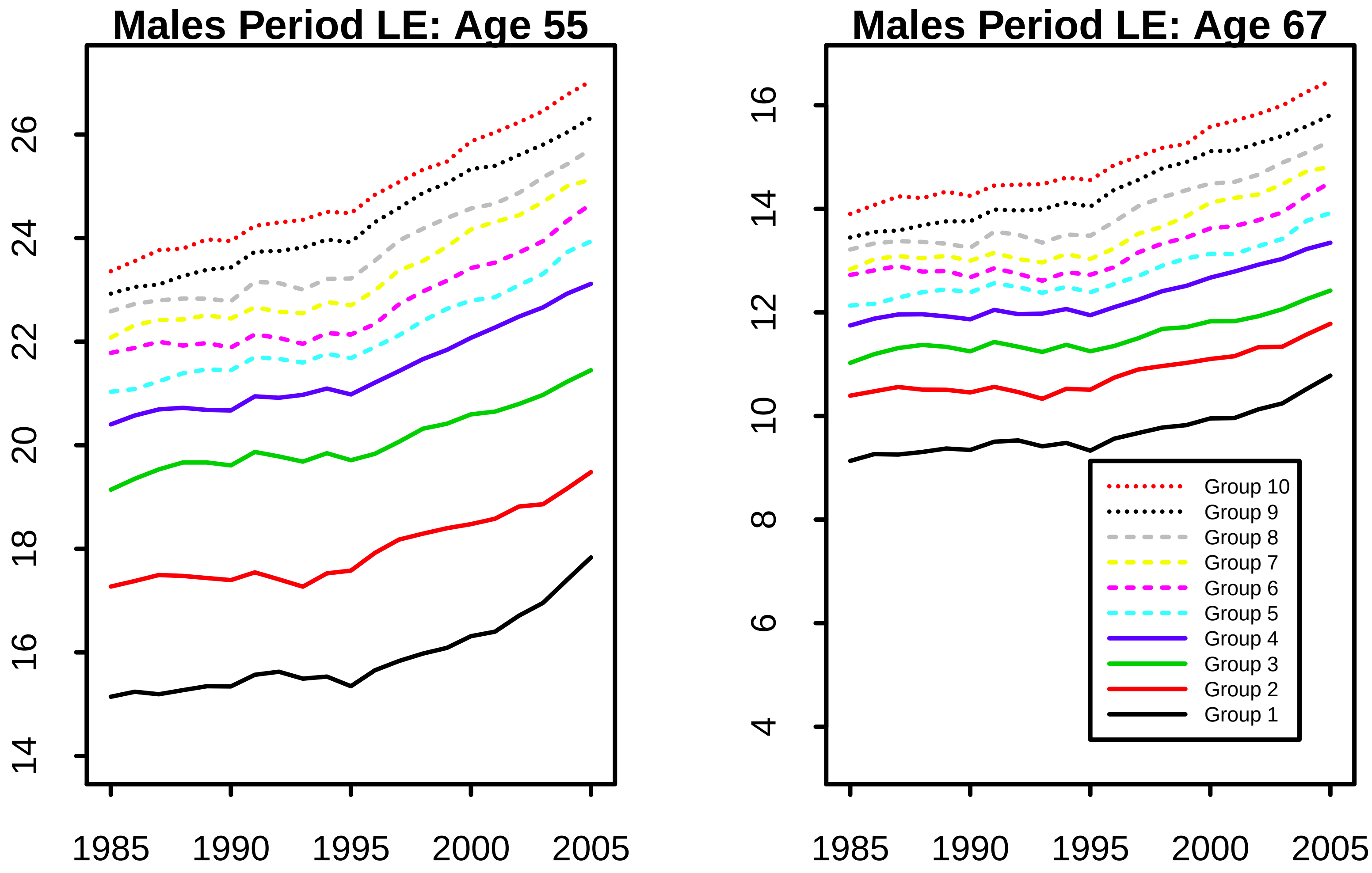
## Time series modelling

- $t \rightarrow t+1$: Allow for correlation
  - between $\kappa_1^{(k)}(t+1)$ and $\kappa_2^{(k)}(t+1)$
  - between groups $k = 1, \ldots, 10$
- Bio-demographical reasonableness

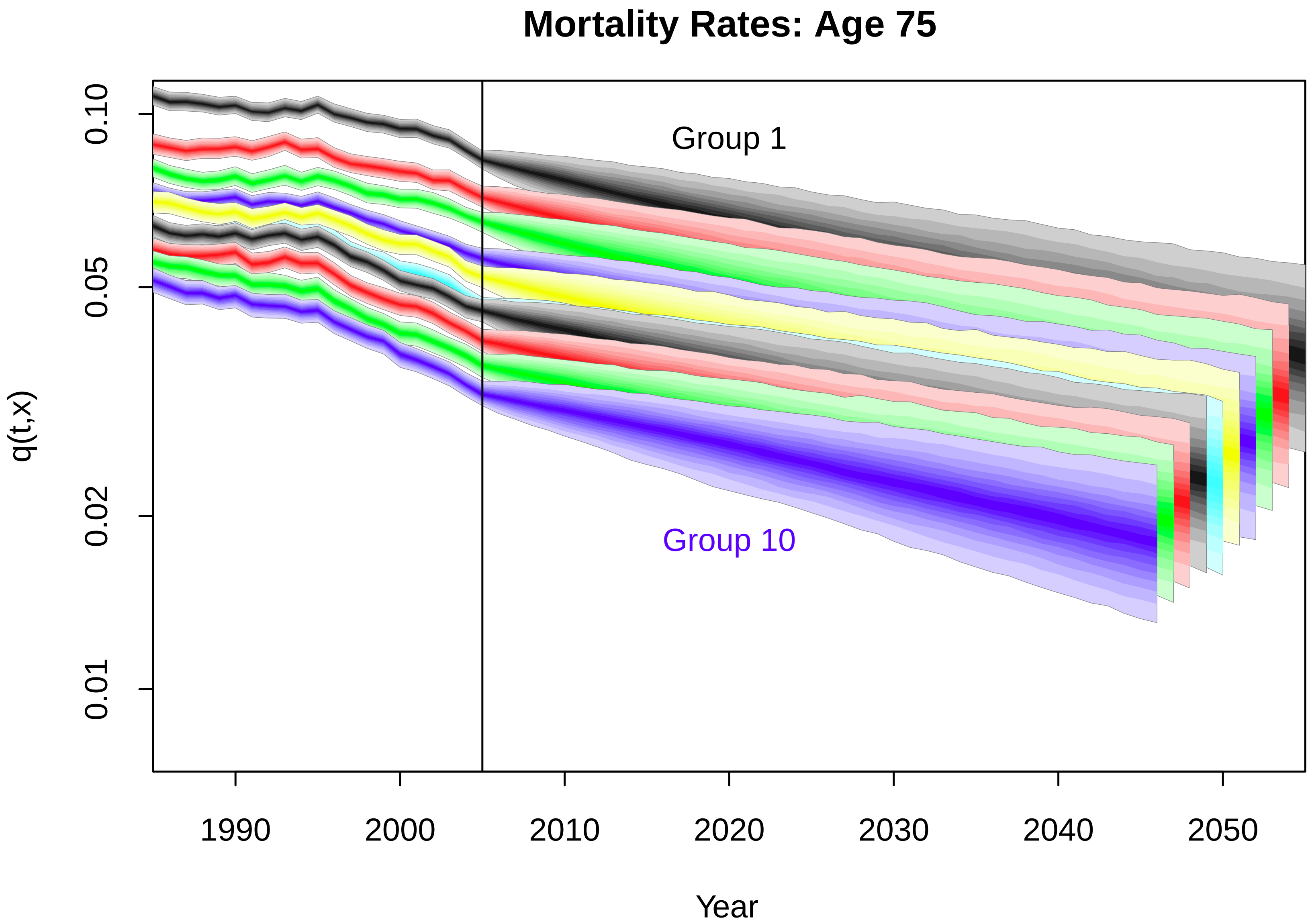
  $\Rightarrow$ key hypothesis: groups should not diverge

  $\Rightarrow$ *group specific* period effects *gravitate* towards the national trend

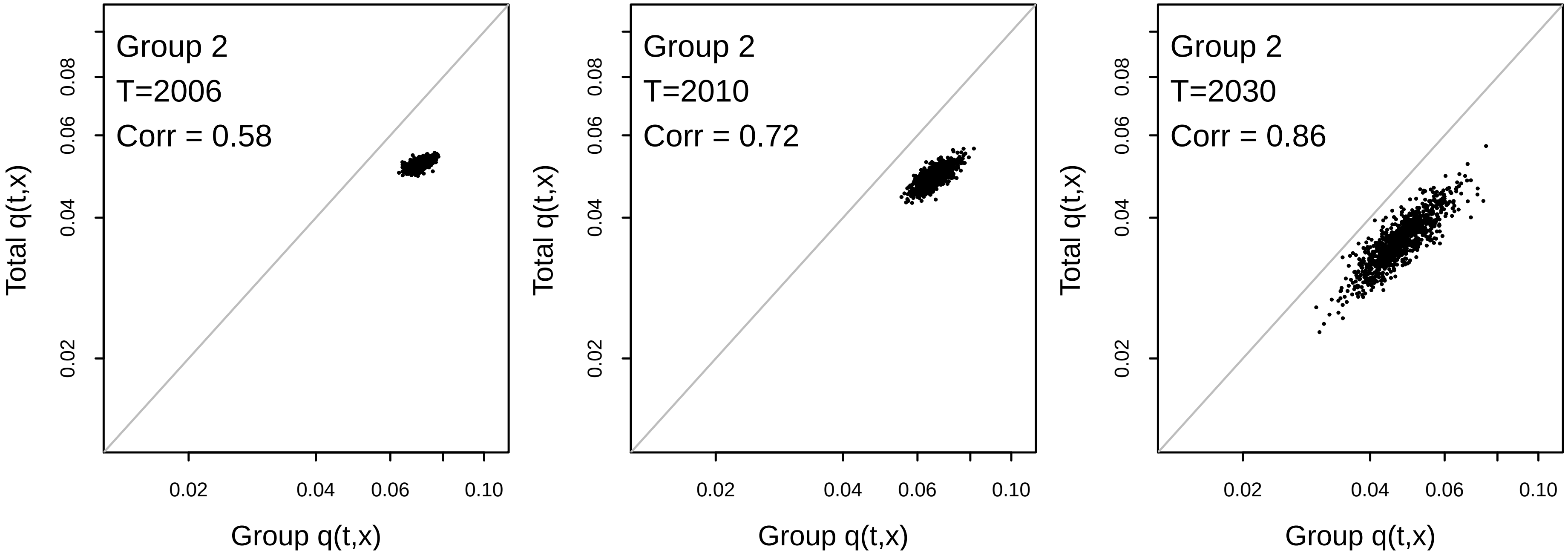
# Life Expectancy for Groups 1 to 10



# Mortality Fan Charts Including Parameter Uncertainty



# Simulated Group versus Population Mortality




As $T$ increases

- Scatterplots become more dispersed
- Shift down and to the left
- Correlation increasess

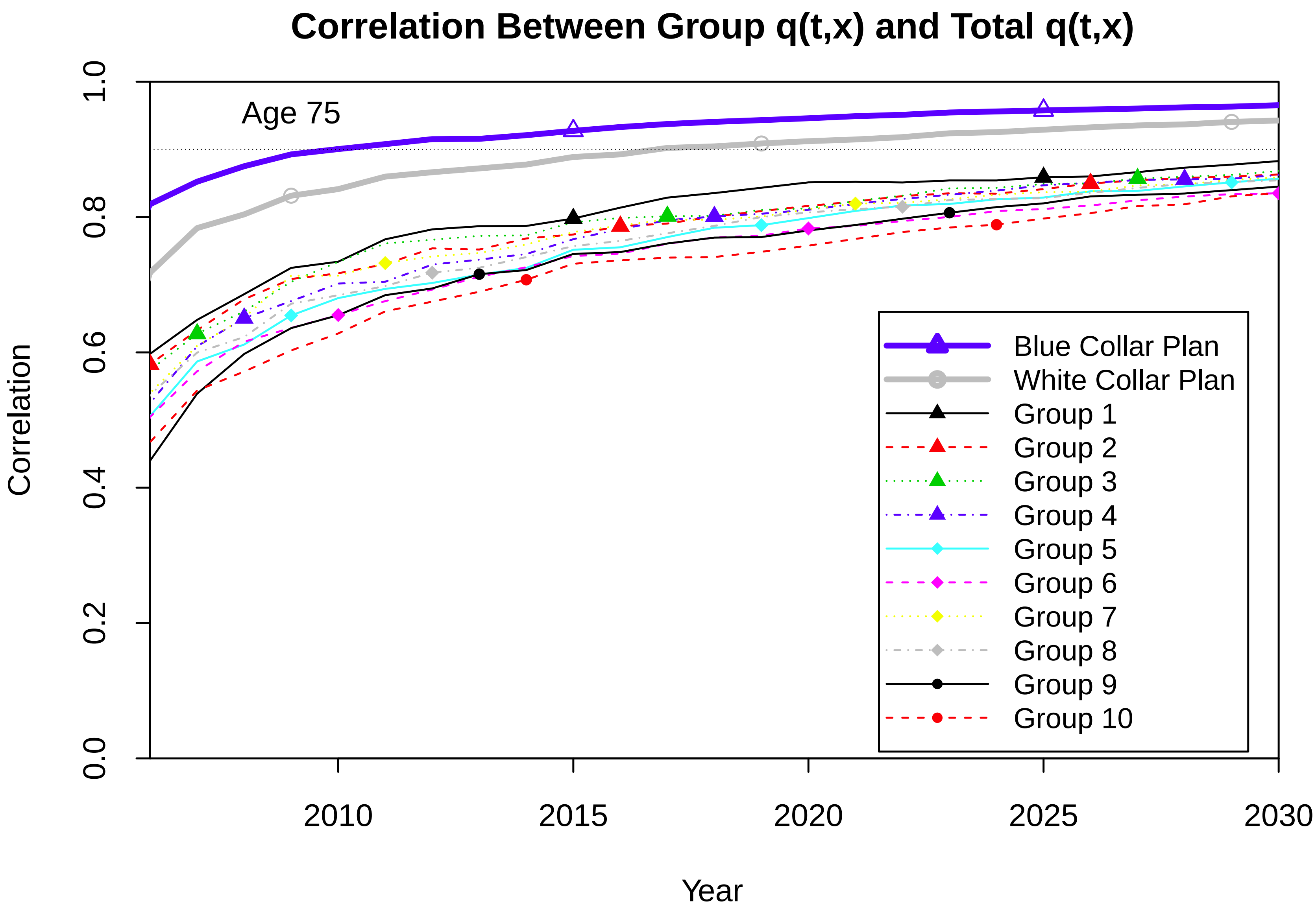
## Forecast Correlations

- Deciles are quite narrow subgroups
- Blue collar pension plan

  $\Rightarrow$ equal proportions of groups 2, 3, 4
- White collar pension plan

  $\Rightarrow$ equal proportions of groups 8, 9, 10

# Forecast Correlations



## Conclusions

- Development of a new multi-population dataset for Denmark

  *strong bio-demographically reasonable group rankings*

  *based on a new measure of affluence*

- Unlike multi-country data

  *a priori* ranking of affluence-related groups

- Proposal for a simple new multi-population model
- Strong correlations over medium to long term
- Correlations depend strongly on diversity of sub-population

# Bonus Slides

# A specific model

$$
\begin{aligned}
\kappa_1^{(i)}(t) &= \kappa_1^{(i)}(t-1) + \mu_1 + Z_{1i}(t) && \text{(random walk)} \\
&\quad -\psi\left(\kappa_1^{(i)}(t-1) - \bar{\kappa}_1(t-1)\right) && \text{(gravity between groups)} \\
\kappa_2^{(i)}(t) &= \kappa_2^{(i)}(t-1) + \mu_2 + Z_{2i}(t) \\
&\quad -\psi\left(\kappa_2^{(i)}(t-1) - \bar{\kappa}_2(t-1)\right)
\end{aligned}
$$

where

$$
\bar{\kappa}_1(t) = \frac{1}{n}\sum_{i=1}^{n} \kappa_1^{(i)}(t) \quad \text{and} \quad \bar{\kappa}_2(t) = \frac{1}{n}\sum_{i=1}^{n} \kappa_2^{(i)}(t)
$$

## A specific model

$$\kappa_1^{(i)}(t) = \kappa_1^{(i)}(t-1) + \mu_1 + Z_{1i}(t) - \psi\left(\kappa_1^{(i)}(t-1) - \bar{\kappa}_1(t-1)\right)$$
$$\kappa_2^{(i)}(t) = \kappa_2^{(i)}(t-1) + \mu_2 + Z_{2i}(t) - \psi\left(\kappa_2^{(i)}(t-1) - \bar{\kappa}_2(t-1)\right)$$

Model structure $\Rightarrow$

- $(\bar{\kappa}_1(t), \bar{\kappa}_2(t)) \sim$ bivariate random walk
- Each $\kappa_1^{(i)}(t) - \bar{\kappa}_1(t) \sim AR(1)$ reverting to $0$
- Each $\kappa_2^{(i)}(t) - \bar{\kappa}_2(t) \sim AR(1)$ reverting to $0$
- $\beta^{(i)}(x)$ vs $\beta^{(j)}(x) \Rightarrow$ intrinsic group differences

## Non-trivial correlation structure: between different ages and groups

$$\begin{aligned}
\kappa_1^{(i)}(t) &= \kappa_1^{(i)}(t-1) + \mu_1 + Z_{1i}(t) - \psi\left(\kappa_1^{(i)}(t-1) - \bar{\kappa}_1(t-1)\right) \\
\kappa_2^{(i)}(t) &= \kappa_2^{(i)}(t-1) + \mu_2 + Z_{2i}(t) - \psi\left(\kappa_2^{(i)}(t-1) - \bar{\kappa}_2(t-1)\right)
\end{aligned}$$

The $Z_{ki}$ are multivariate normal, mean 0 and

$$Cov(Z_{ki}, Z_{lj}) = \begin{cases} v_{kl} & \text{for } i = j \\ \rho v_{kl} & \text{for } i \neq j \end{cases}$$

$\rho =$ cond. correlation between $\kappa_1^{(i)}(t)$ and $\kappa_1^{(j)}(t)$ etc.

## Comments

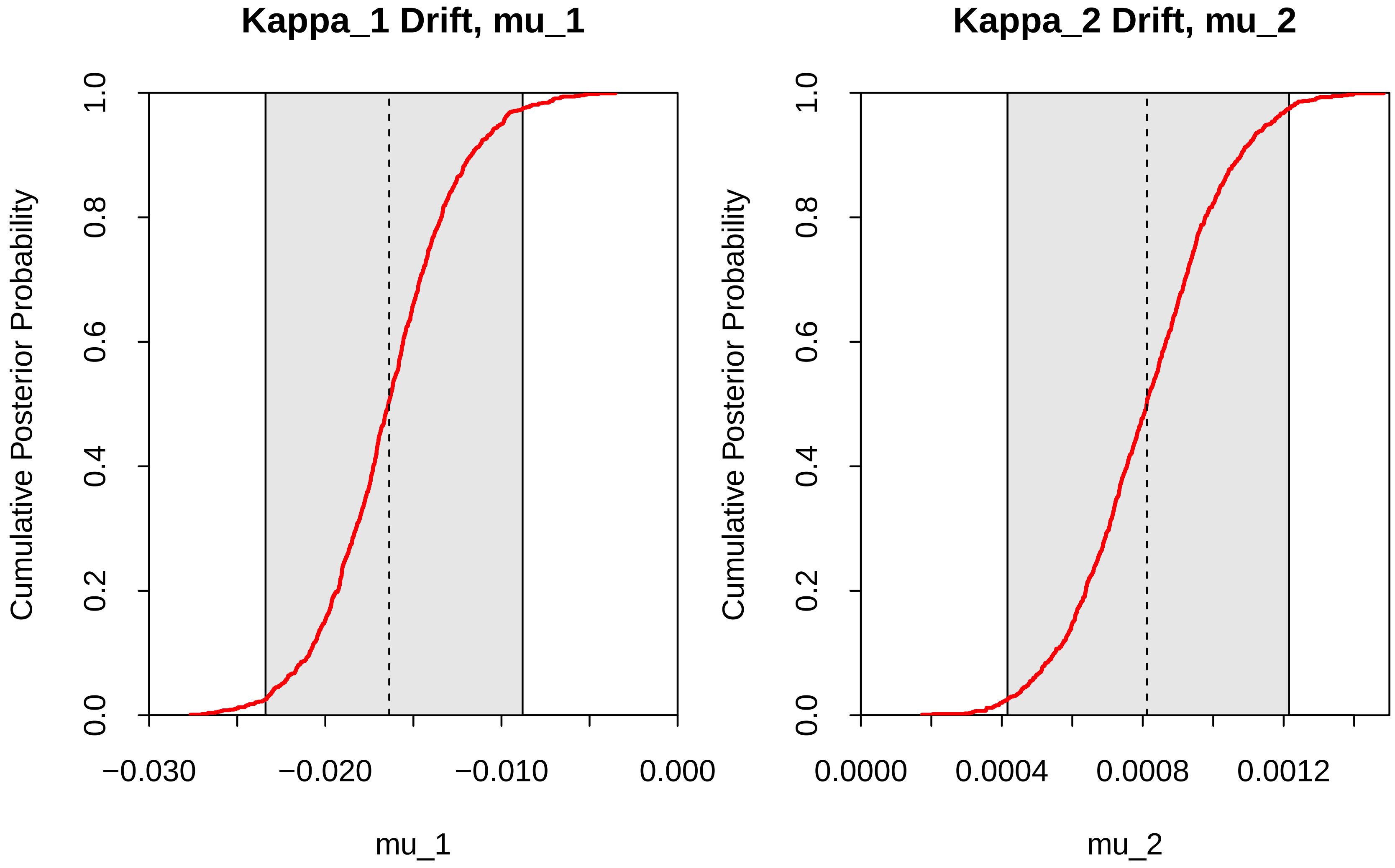
- Model is very simple
  - One gravity parameter, $0 < \psi < 1$
  - One between-group correlation parameter, $0 < \rho < 1$
- Many generalisations are possible
- But more parameters $+$ more complex computing
- This simple model seems to fit quite well.
- Nevertheless $\Rightarrow$ work in progress

## Prior distributions

- As uninformative as possible
- $\mu_1, \ \mu_2 \sim$ improper uniform prior
- $\{v_{ij}\} \sim$ Inverse Wishart
- $\rho \sim \text{Beta}(2, 2)$
- $\psi \sim \text{Beta}(2, 2)$

State variables and process parameters estimated using MCMC (Gibbs + Metropolis-Hastings)

## Posterior Distributions and $95\%$ Credibility Intervals




Note: $-\mu_1 =$ global improvement rate

# Posterior Distributions and $95\%$ Credibility Intervals