UK Research and Innovation

THE UNIVERSITY of EDINBURGH

ACRC
Advanced Care Research Centre

Nonparametric Bayesian Multi-facet Clustering for Longitudinal Data

Luwei Wang, Kieran Richards, Sohan Seth

Data Science Unit, School of Informatics, University of Edinburgh

Biomedical AI
UKRI CENTRE FOR DOCTORAL TRAINING

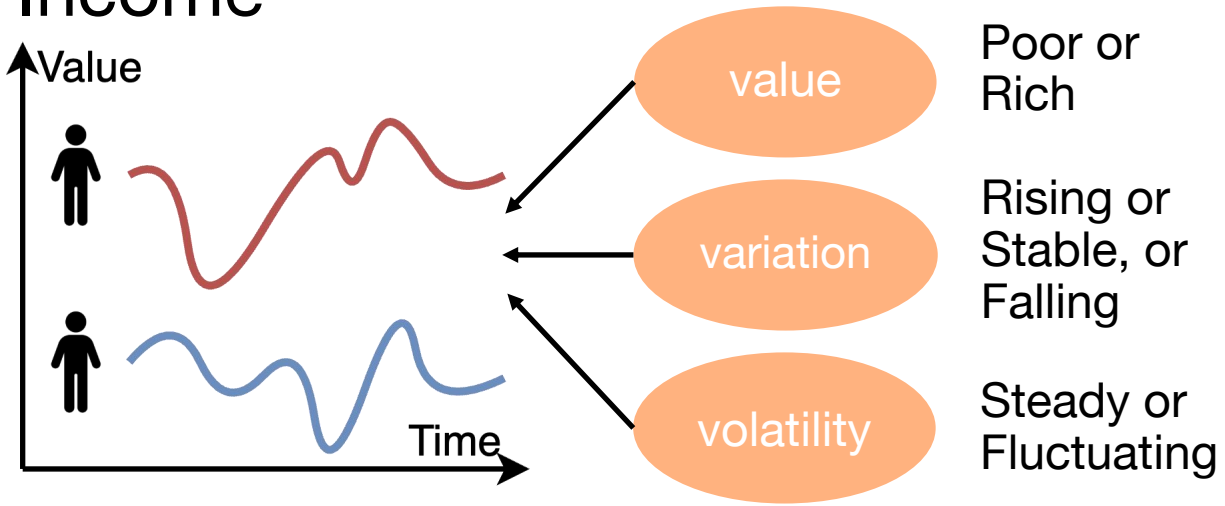
UKAIRS 2025

AI MULTIPLY

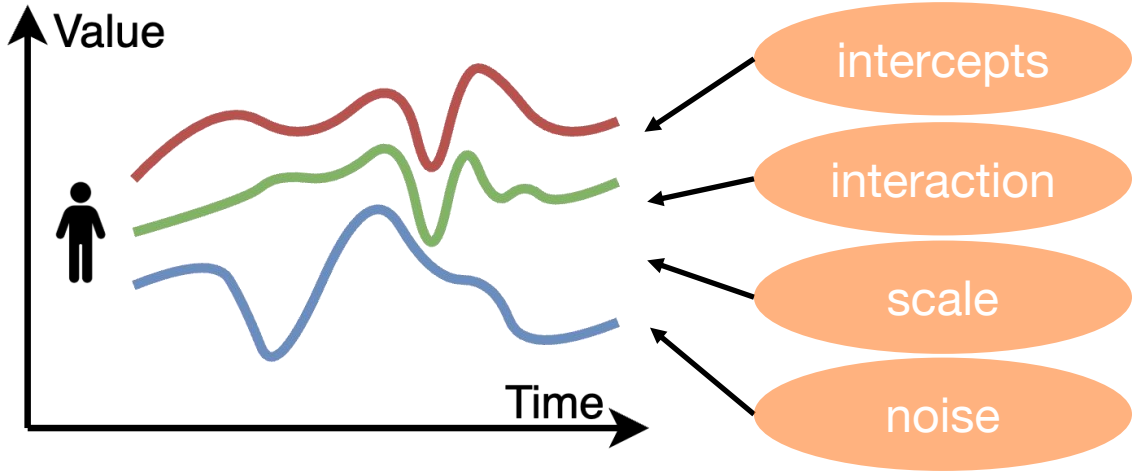
AIM CISC

Multiple facets of longitudinal data

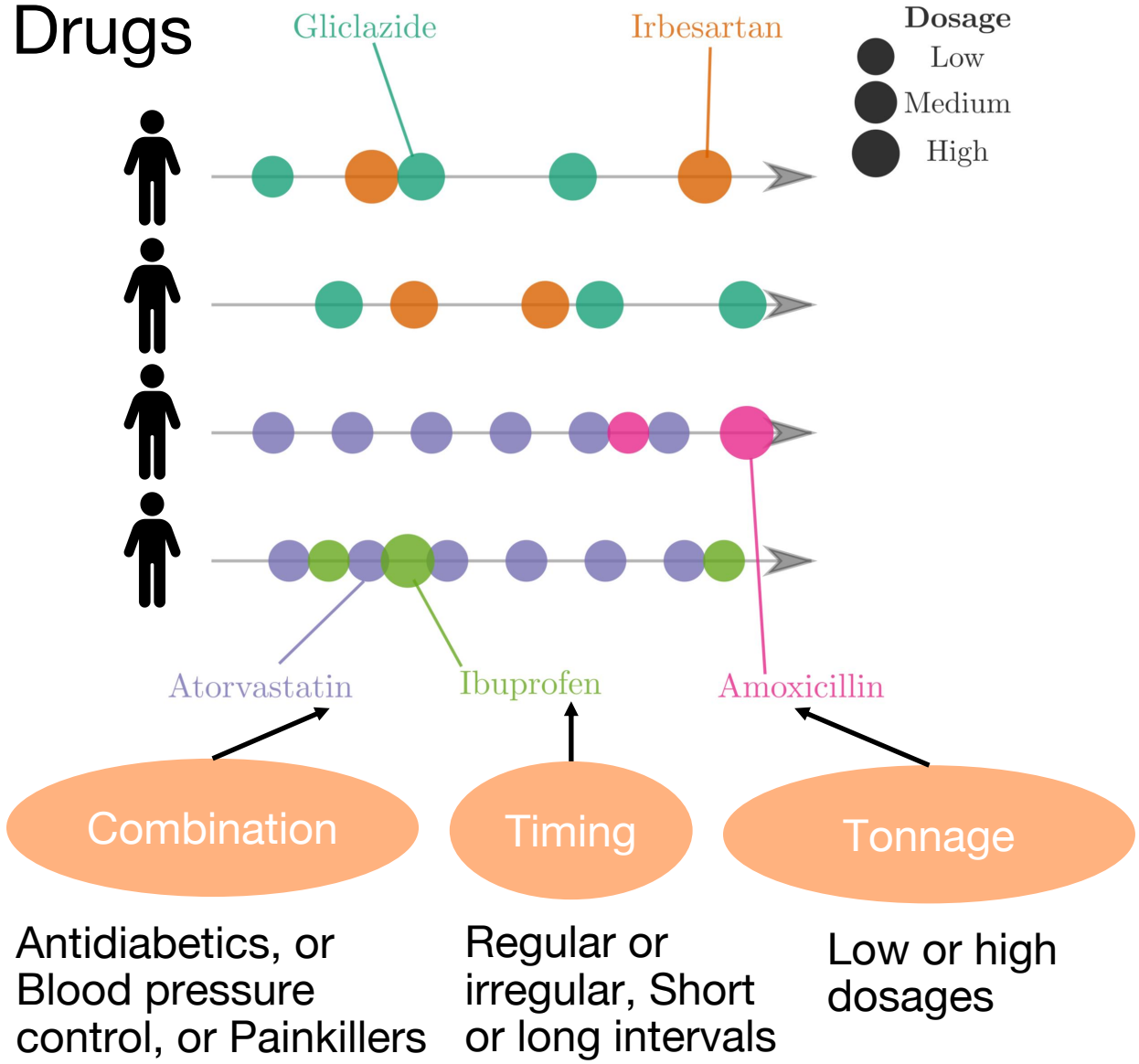
Income



Frailty, wellbeing & social isolation

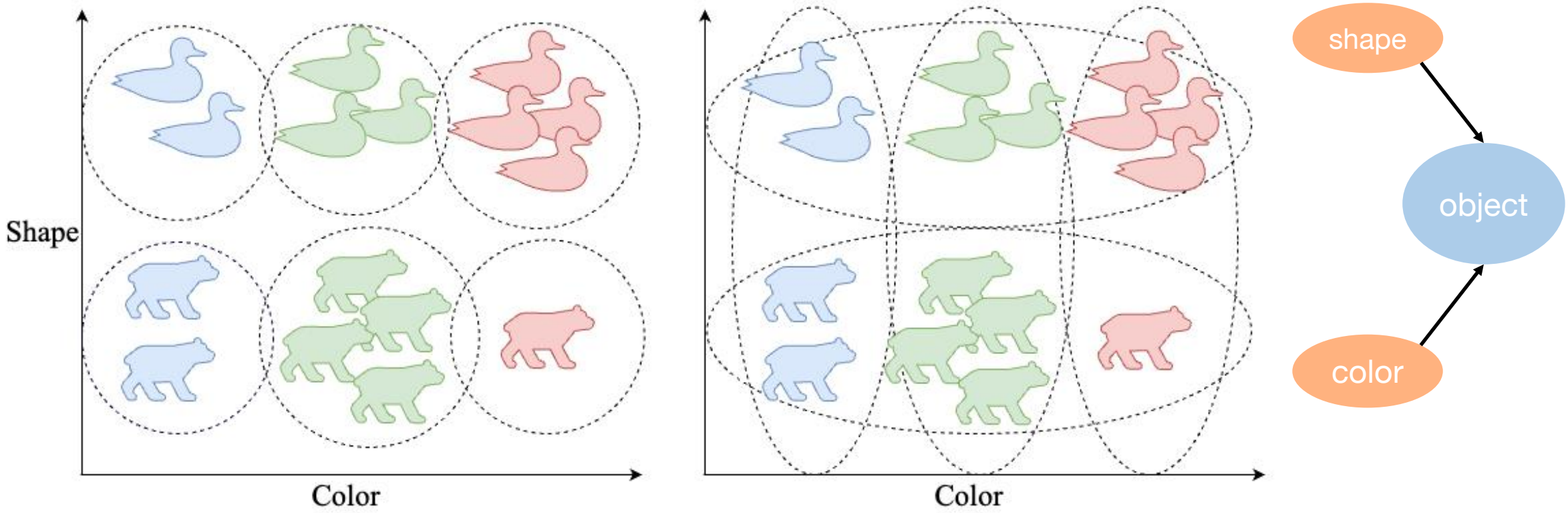


Drugs



Disentangle facets for interpretability

- **Traditional clustering** aggregates all facets/characteristics
  → poor interpretability
- **Multi-facet clustering:** along different facets/characteristics



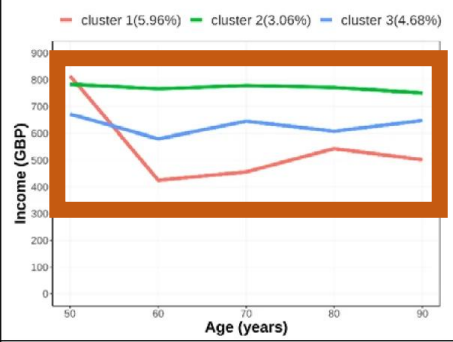
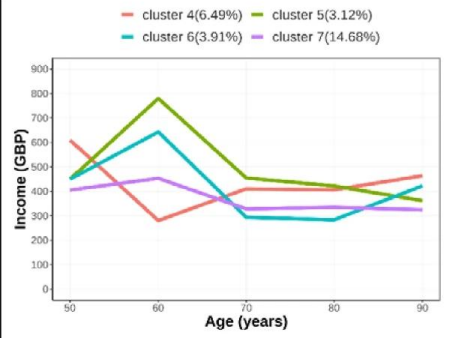
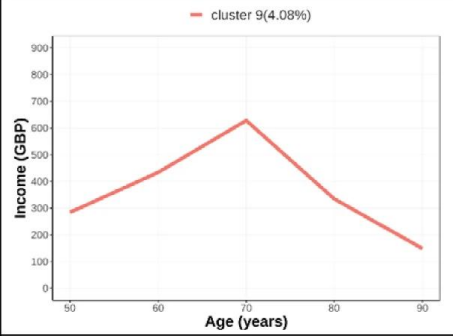
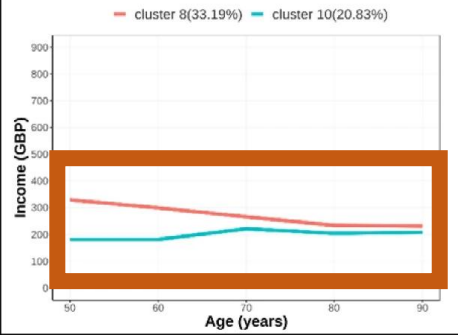
Single-facet versus multi-facet clustering.

Method

- **Multi-Facet Mixture Model (MMM):** separate clustering per facet

$$\sum_{k_1=1}^{K_1} \cdots \sum_{k_F=1}^{K_F} \pi_{k_1}^{(1)} \cdots \pi_{k_F}^{(F)} p(\mathbf{y} \mid \boldsymbol{\theta}_{k_1}^{(1)}, \ldots, \boldsymbol{\theta}_{k_F}^{(F)})$$

- Independent facets apriori
- Time series data i.e. y(t)
- Facets defined by users
- *Dirichlet Process prior* for automatic cluster discovery
- *Variational Bayesian inference* for scalability vs MCMC
- Implemented for two time series models
  - Nonlinear Growth Model (NLG) & Vector Autoregressive Model (VAR)
  - More being added!

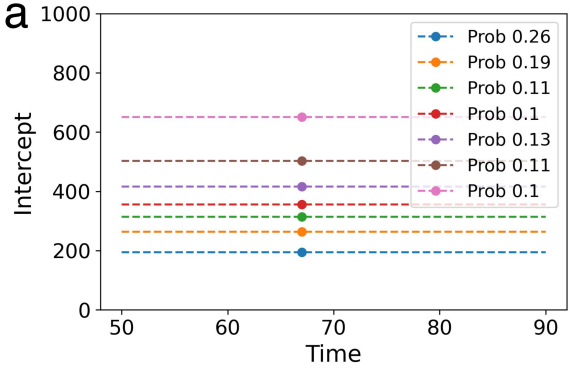
Income trajectory clusters	Income trajectory group (combines clusters)
	**Luxury:** These trajectories comprise 13.7% of the sample and share a high income that is at or above £500 per week for most of retirement. Cluster 1 (Luxury – pre-retirement income drop) – 5.7% (N=775) Clusters 2 (Luxury, highest stable income) - 3.1% (N=398) Cluster 3 (Luxury – medium stable income) – 4.7% (N=609)
	**Comfortable:** These clusters comprise 28% of the population and share a mid-level income from age 65 that is between £300-£500 per week. Cluster 4 (Comfortable: pre-retirement drop in income) – 6.5% (N=844) Cluster 5 (Comfortable: spike in pre-retirement income) – 3.1% (N=406) Cluster 6 (Comfortable: spike in pre-retirement income) – 3.9% (N=508) Cluster 7 (Comfortable – stable income trajectory) – 14.7% (N=1,909)
	**Boom-to-Bust** This group comprises 4.1% (N=530) of the sample and one cluster (9) with a spike of very high income at age 70 of £600 per week falling below £300 per week at the oldest ages (80+).
	**Always Poor:** These trajectories share an income that is typically below £300 per week from age 65 and to around £230 from age 80. Clusters 8 (Always Poor: higher within-group income) – 33% (N=4,317) Cluster 10 (Always Poor: lower within group income) 21% (N=2,709)

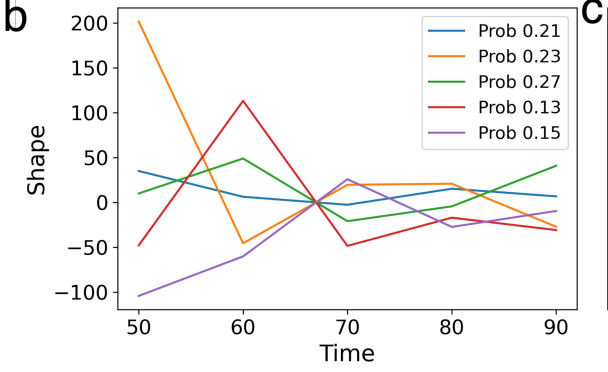
Single-facet clustering

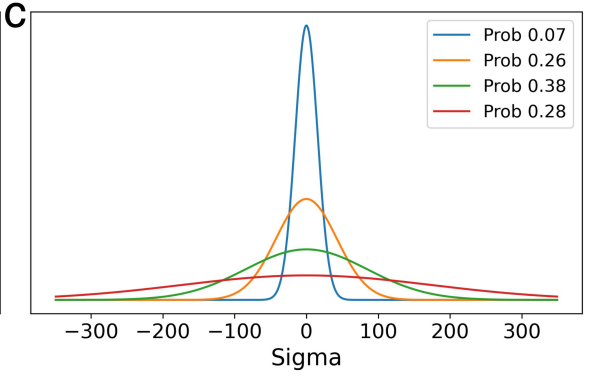
- **English Longitudinal Study of Ageing (ELSA)**
  - N = 13,002
  - Age: 50-90
  - Missingness: 86.8%
- **10** income trajectory clusters
- Manually grouped to **4 super-clusters** based on income at retirement
  - Luxury
  - Comfortable
  - Boom-to-bust
  - Always Poor
- **Similar variations** of trajectory
- **No volatility**

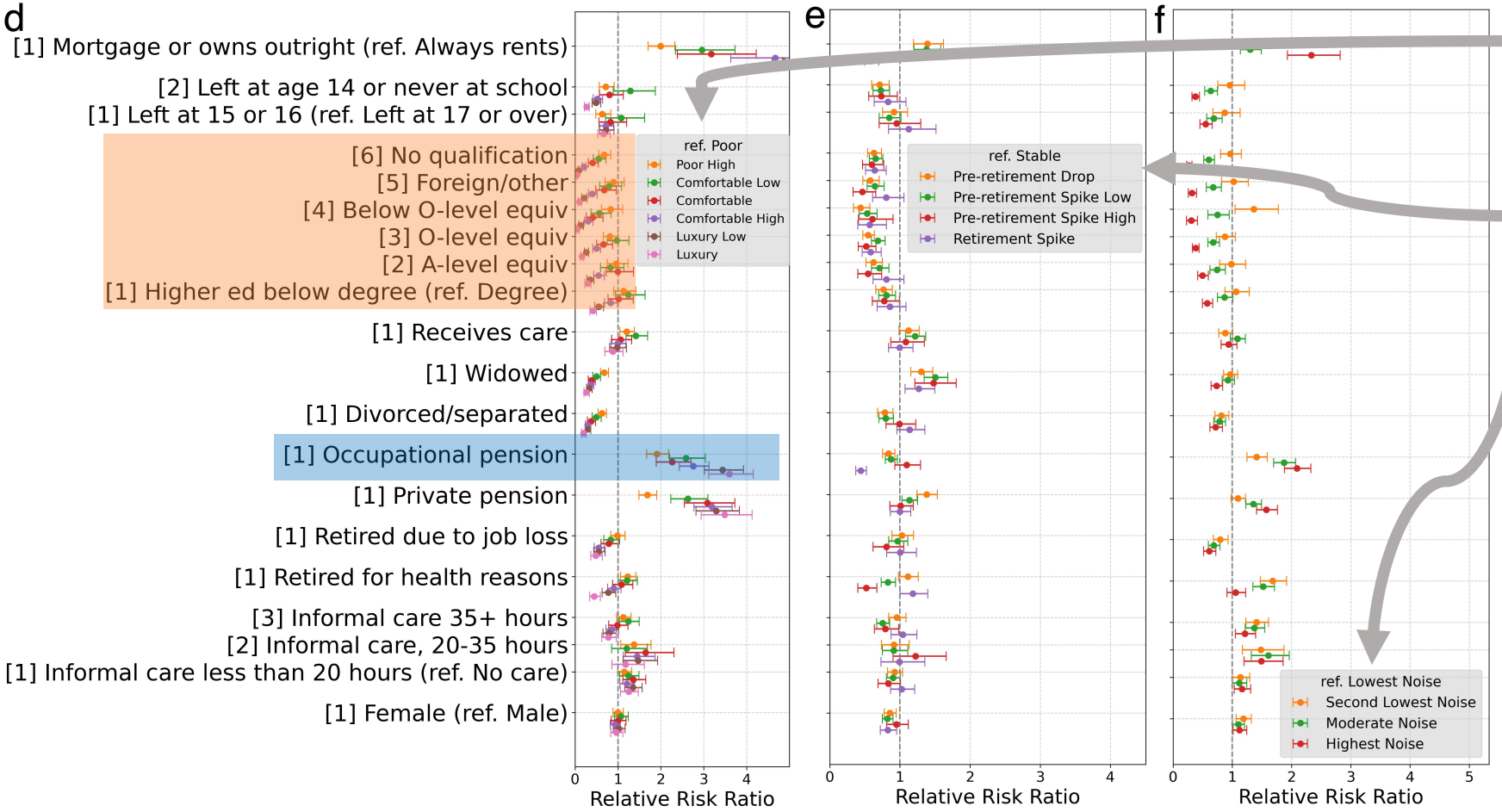
Marshall, A., Eke, C., Guthrie, B. et al. Income Trajectories and Precarity in Later life. Population Ageing 17, 335–363 (2024). https://doi.org/10.1007/s12062-023-09437-2

Income clusters & their drivers









7 value clusters

(Poor → Luxury)

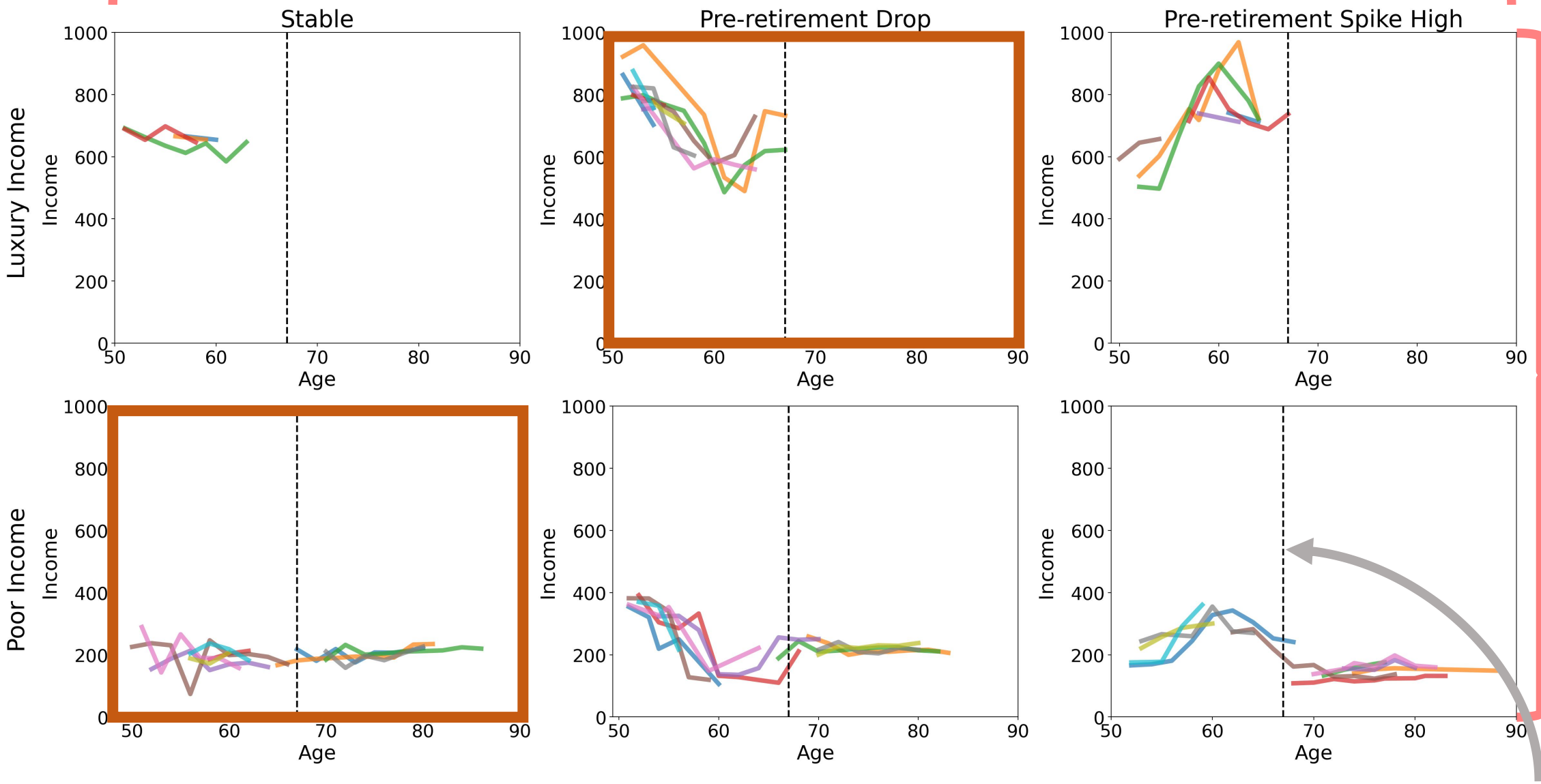
5 variation clusters

4 volatility clusters

Low education ↓ luxury

Pensions ↑ luxury

Facet 2: Income variation



Facet 1: Income value

retirement age at 67

Summary

Real-world data are multi-faceted

1. **Interpretable** by disentangling multiple facets
2. Current implementation supports
   a) Large datasets
   b) Automatic number of clusters
3. Broader **applicability** to various trajectories
   a) Socioeconomics
   b) Health
   c) and many more!

Thank you and get in touch!

Full Paper

MMM Code