Bayesian Local Clustering of Age-Period Mortality Surfaces across Multiple Countries

Giovanni Romanò,¹ Emanuele Aliverti² and Daniele Durante³

¹Department of Decision Sciences, Bocconi University, Italy, ²Department of Statistical Sciences, University of Padova, Italy, and ³Department of Decision Sciences and Bocconi Institute for Data Science and Analytics, Bocconi University, Italy

Address for correspondence: Giovanni Romanò, Department of Decision Sciences, Bocconi University, Via Roentgen 1, Milan, Italy. Email: giovanni.romano4@phd.unibocconi.it

Abstract

Although traditional literature on mortality modeling has focused on single countries in isolation, recent contributions have progressively moved toward joint models for multiple countries. Besides favoring borrowing of information to improve age-period forecasts, this perspective has also potentials to infer local similarities among countries' mortality patterns in specific age classes and periods that could unveil unexplored demographic trends, while guiding the design of targeted policies. Advancements along this latter relevant direction are currently undermined by the lack of a multi-country model capable of incorporating the core structures of age-period mortality surfaces together with clustering patterns among countries that are not global, but rather vary locally across different combinations of ages and periods. We cover this gap by developing a novel Bayesian model for log-mortality rates that characterizes the age structure of mortality through a B-spline expansion whose countryspecific dynamic coefficients encode both changes of this age structure across periods and also local clustering patterns among countries under a time-dependent random partition prior for these country-specific dynamic coefficients. While flexible, this formulation admits tractable posterior inference leveraging a suitably-designed Gibbs-sampler. The application to mortality data from 14 countries unveils local similarities highlighting both previously-recognized demographic phenomena and also yet-unexplored trends.

Key words: B-splines, dynamic clustering, mortality rates, random partition model

1. Introduction

The changes in life expectancy, population structure and welfare systems over the past decades have stimulated a growing demand for novel statistical models capable of characterizing heterogenous mortality patterns across ages, periods and countries, while providing reliable probabilistic forecasts with rigorous uncertainty quantification [e.g., Lutz and Kc, 2010, Raftery et al., 2013, Hunt and Blake, 2021]. Advancements along these lines are fundamental in guiding health-care, social, environmental and retirement policies, thereby motivating active research on mortality modeling within several fields, such as demography [e.g., Lee and Miller, 2001, Li and Lee, 2005, De Jong and Tickle, 2006, Raftery et al., 2013, Li et al., 2013, Hyndman et al., 2013, Mazzuco et al., 2018, Camarda, 2019, Léger and Mazzuco, 2021], statistics [e.g., Lee and Carter, 1992, Hyndman and Ullah, 2007, Alexopoulos et al., 2019, Tang et al., 2022, Aliverti et al., 2022, Lam and Wang, 2023, Pavone et al., 2024, Debon et al., 2023, Dimai, 2025] and actuarial sciences [e.g., Haberman and Renshaw, 2011, Hatzopoulos and Haberman, 2013, Kleinow, 2015, Antonio et al., 2015, Currie, 2016, Enchev et al., 2017, Wong et al., 2018, Dong et al., 2020, Scognamiglio, 2022], among others. As clarified in these contributions, such an active research on mortality has witnessed in the recent years a progressive shift away from analyzing the single countries in isolation and towards joint modeling of age-period mortality surfaces within a multi-country setting.

The above perspective is inherently motivated by the fact that countries are not isolated entities, but rather display similar global trends in age-specific mortality patterns (see Figure 1) regulated by common exogenous and endogenous factors, such as advancements in health-care [e.g., Vallin and Meslé, 2004]. As such, joint modeling of multiple countries can facilitate effective borrowing of information to improve age-period mortality forecasts, while opening the avenues for inference on local similarities and differences among countries' mortality patterns in specific age classes and periods that could unveil unexplored demographic trends, possibly arising from targeted policies adopted by certain countries. Addressing both objectives would require a unique formulation that accounts not only for the core structures of country-specific age-period mortality surfaces, but also for local clustering patterns between countries arising from overlaps among these country-specific surfaces in particular combinations of ages and periods. Although state-of-theart multi-country mortality models are still not designed to include these fundamental dynamics,



Figure 1: Log-mortality rates at age 0 (first panel) and 55 (second panel) in Italy (ITA), Sweden (SWE), the United Kingdom (UK) and the United States (USA) between 1933 and 2020.

local heterogeneities in country-specific mortality rates across ages and periods are recognized in demographic research [e.g., Vallin and Meslé, 2004, Vaupel et al., 2011] and find strong evidence in mortality data. For example, the first panel of Figure 1 illustrates the infant log-mortality rates in Italy, Sweden, the United Kingdom (UK) and the United States (USA) from 1933 to 2020, as retrieved from the Human Mortality Database (HMD: https://www.mortality.org/). During the first forty years, the UK and the USA display overlapping mortality trajectories, which then diverge after 1980. Conversely, Italy follows a distinct path until 1985, after which it aligns with the UK for approximately ten years. Interestingly, as shown within the second panel of Figure 1, these grouping structures vary not only across periods, but also with different ages. For instance, the clustering among the UK and the USA at infant ages is no more visible for individuals aged 55, who show, instead, a remarkable overlap between Italy and the UK for a large time window.

Accounting for the above patterns via a principled model-based representation would not only provide a more realistic characterization of age-period mortality surfaces across multiple countries, but could also open the avenues for rigorously answering important scientific questions, e.g., on the differences among low-mortality countries [Oeppen and Vaupel, 2002, Vaupel et al., 2011] and the corresponding socio-economic determinants [Marmot, 2005]. As discussed previously, although the literature on multi-country mortality modeling has witnessed extensive advancements over the recent years, state-of-the-art contributions are not designed to address this endeavor. In fact, a common solution in multi-country mortality modeling relies on extending structured bilinear decompositions of age-period mortality surfaces for a single country, such as the one by Lee and Carter [1992], to joint formulations that include country-specific parameters and a common component shared across periods [e.g., Li and Lee, 2005] or ages [e.g., Kleinow, 2015]. This general perspective improves forecasts and can be further extended to more sophisticated representations leveraging multi-country generalizations [Hyndman et al., 2013, Lam and Wang, 2023] of functional principal components constructions [Hyndman and Ullah, 2007], or joint decompositions of the country-age-period mortality tensor [Dong et al., 2020]; see also Enchev et al. [2017]. However, all these solutions are not designed to infer group structures among countries based on similarities in the associated mortality rates. Such an issue is also found in related Bayesian hierarchical formulations relying on conditionally independent models for each country linked by a common prior distribution on shared underlying parameters [e.g., Raftery et al., 2013, Antonio et al., 2015, Aliverti et al., 2022].

An effective direction for overcoming the aforementioned issues is to move towards more recent extensions of the above formulations which explicitly include notions of clustering among countries in terms of the corresponding mortality patterns [Hatzopoulos and Haberman, 2013, Léger and Mazzuco, 2021, Schnürch et al., 2021, Tang et al., 2022, Scognamiglio, 2022, Debon et al., 2023, Dimai, 2025]. Albeit providing meaningful representations for model-based clustering of mortality surfaces, the overarching focus of these extensions is on global grouping structures, instead of local ones changing with different combinations of ages and periods. As such, countries are not allowed to display varying clustering behaviors at different ages [Léger and Mazzuco, 2021], time periods [Hatzopoulos and Haberman, 2013], or both dimensions [Tang et al., 2022, Schnürch et al., 2021, Scognamiglio, 2022, Dimai, 2025]. Recalling the previous discussion of Figure 1, these constraints

are not supported by the observed age-period mortality data and hinder the possibility of learning nuanced grouping structures, possibly informing on unexplored localized demographic trends that are visible only for specific combinations of ages and periods.

To our knowledge, the only attempt to remove the constraints imposed by the above global clustering perspectives can be found in the application of latent class clustering models to multicountry age-period mortality data in Debon et al. [2023]. Although this contribution has the merit of recognizing the importance of moving beyond global clustering perspectives, a direct application of latent class clustering methods without the additional inclusion of the specific structures of age-period mortality surfaces could undermine the flexibility of the resulting procedure in uncovering local clustering patterns. This potential issue finds evidence in Figure 1 of Debon et al. [2023] where the inferred grouping structures resemble more closely those obtained under a global perspective, than a local one. For example, according to Figure 1 of Debon et al. [2023], at infant ages, all countries share the same cluster in the entire time window analyzed, a behavior which is not in line with the local clustering structures displayed by the observed mortality data in our Figure 1. In addition, similarly to all cluster-based extensions of multi-country mortality models, also Debon et al. [2023] requires to specify the unknown number of groups, a challenging task in practice, without a uniquely-accepted solution.

Motivated by the above discussion and by the impact of addressing the aforementioned challenges in multi-country mortality modeling, we propose and develop in Section 2 an innovative and principled Bayesian formulation that accounts for the core structures of age-period mortality surfaces, and crucially incorporates clustering patterns among countries which are allowed to vary flexibly across different combinations of ages and periods. This is accomplished by modeling the smooth age patterns of mortality via a flexible linear combination of B-spline bases [e.g., Camarda, 2019, Pavone et al., 2024, whose country-specific dynamic coefficients evolve across calendar years through carefully-designed stochastic processes of time relying on a temporal random partition prior inspired by the general construction in Page et al. [2022]. Crucially, these stochastic processes for the joint time trajectories of the country-specific coefficients associated to the different B-spline bases provide a principled characterization of the time changes in the age patterns of mortality across periods, while allowing the grouping structures exhibited by countries to change both in time and across the bases' coefficients associated with the different ages. As such, the resulting representation facilitates the identification of locally converging or diverging trends in multi-country age-period mortality surfaces, while crucially preserving a structured representation that accounts for the core characteristics of these surfaces.

As clarified in Section 3, the proposed structured representation, albeit flexible, is amenable to tractable posterior inference via a carefully-designed Gibbs-sampling algorithm that learns automatically the unknown total number of clusters and facilitates both point estimation and uncertainty quantification on mortality patterns and grouping structures. The simulation studies within Section 4 and the application to mortality data of 14 countries from 1933 until 2020 in Section 5, not only confirm the ability of the proposed model to accurately learn these local grouping structures, but also unveil unique localized similarities among specific countries, which highlight both known demographic phenomena and also yet-unexplored trends acting on selected countries over

specific combinations of ages and periods. Concluding remarks can be found in Section 6 where we also clarify that, although motivated by multi-country age-period mortality data, the proposed model has broader scope and impact, in that it allows to detect localized overlaps among surfaces associated with different populations. To our knowledge, general methodological contributions exploring this direction are limited.

2. Model Formulation

Let us denote with d_{ixt} and E_{ixt} , respectively, the observed death counts and the average number of individuals at risk within country i = 1, ..., n, at age $x \in \mathcal{X}$ and for period t = 1, ..., T (corresponding, in our case, to calendar years). Consistent with the overarching focus in mortality models for both single and multiple countries [see, for example, Currie, 2016, Enchev et al., 2017, Hunt and Blake, 2021], our interest lies in the analysis of the country-specific age-period log-mortality rates log $m_{ixt} = \log(d_{ixt}/E_{ixt})$ for which we assume

$$\log m_{ixt} = f_{it}(x) + \varepsilon_{ixt}, \quad \text{with} \quad \varepsilon_{ixt} \sim \mathcal{N}(0, \sigma_i^2), \tag{1}$$

independently for $i = 1, ..., n, x \in \mathcal{X}$ and t = 1, ..., T, where $f_{it}(x) = \mathbb{E}[\log m_{ixt} | f_{it}(x)]$ is the expected log-mortality rate surface for the *i*-th country expressed as a function of age $x \in \mathcal{X}$ that is allowed to vary across periods t = 1, ..., T. The general surface plus Gaussian noise formulation in (1) is common to several mortality models for both single and multiple countries [e.g., Currie, 2016, Enchev et al., 2017, Hunt and Blake, 2021]. In addition, as recently proved by Pavone et al. [2024, Proposition 2.1] under a single-country perspective, when E_{ixt} is sufficiently large (as in standard demographic settings), the above formulation arises directly from an underlying Poisson log-normal model for the observed death counts d_{ixt} that properly accounts for possible overdispersion; see also Wong et al. [2018]. Nonetheless, as discussed in Section 1, none of the available multi-country models provides a structured representation for $f_{it}(x)$ that incorporates the core age-period patterns of mortality, while allowing countries to cluster differently as x and t vary.

Addressing the above gap would require the design of a structured representation for the expected log-mortality rate surface which ensures that $f_{it}(x)$ varies (i) smoothly as a function of age x, for every i = 1, ..., n and t = 1, ..., T, and (ii) dynamically with periods t, for each i = 1, ..., n and $x \in \mathcal{X}$, while (iii) allowing $f_{it}(x)$ and $f_{i't}(x)$ for any two generic countries i and i' to cluster (i.e., display similar values) only for those combinations of x and t in which there is empirical evidence of overlapping patterns in the log-mortality rates of i and i'. To this end, a natural option for addressing objective (i) in a way that facilitates also inclusion of (ii) and (iii) is to represent $f_{it}(x)$ through the B-spline expansion

$$f_{it}(x) = \sum_{j=1}^{p} \beta_{ijt} g_j(x), \qquad i = 1, \dots, n, \ x \in \mathcal{X}, \ t = 1, \dots, T,$$
(2)

where $[g_1(x), \ldots, g_p(x)]$ denotes a set of common B-splines bases [e.g., Eilers and Marx, 1996] with associated coefficients $[\beta_{i1t}, \ldots, \beta_{ipt}]$ that are allowed to vary with countries $i = 1, \ldots, n$ and periods $t = 1, \ldots, T$. Although related representations have been mostly explored in separate analyses of single countries in isolation [e.g., Currie et al., 2004, Camarda, 2019, Pavone et al., 2024], the expansion in (2) provides an effective construction in our multi-country setting which ensures that $f_{it}(x)$ is a smooth function of age, for every i = 1, ..., n and t = 1, ..., T, whose dynamic changes across periods and local clustering patterns among countries can be regulated by a finite set of coefficients $[\beta_{i1t}, \ldots, \beta_{ipt}]$, for i = 1, ..., n and t = 1, ..., T. As such, objective (*i*) is addressed by construction, while (*ii*)–(*iii*) can be accomplished by allowing time changes and ties in these coefficients, respectively. In particular, notice that, when $\beta_{ijt} = \beta_{i'jt}$, then countries *i* and *i'* display under (2) similar mortality rates in period *t* for the age interval associated with basis $g_i(x)$.

To formalize this idea, let $c_{ijt} \in \{1, \ldots, K_{jt}\}$ be the cluster membership indicator for country *i*, with respect to the *j*-th basis in period *t*, and denote with $\mathbf{c}_{jt} = [c_{1jt}, \ldots, c_{njt}] \in \{1, \ldots, K_{jt}\}^n$ the vector comprising the memberships for the *n* countries, for every $j = 1, \ldots, p$ and $t = 1, \ldots, T$, where $K_{jt} \leq n$ is the total number of clusters at the pair (j, t). Then, ties among the coefficients $\beta_{1jt}, \ldots, \beta_{njt}$ can be readily incorporated, for every $j = 1, \ldots, p$ and $t = 1, \ldots, T$, by letting

$$\beta_{ijt} = \beta^*_{c_{ijt}jt}, \quad \text{for all } i = 1, \dots, n, \ j = 1, \dots, p, \ t = 1, \dots, T,$$
(3)

where $\beta_{c_{ijt}jt}^{\star} \in \mathbb{R}$ is the value of the *j*-th B-spline coefficient in period *t* associated with the cluster to which country *i* has been allocated. As such, all countries belonging to the same generic cluster *k* (for the *j*-th spline basis within period *t*) will display the same value β_{kjt}^{\star} for the associated coefficient, thereby addressing *(iii)*. Crucially, \mathbf{c}_{jt} varies with both *j* and *t*, and therefore, any generic pair countries *i* and *i'* is allowed to cluster locally only for specific ages and periods, consistent with the original motivations underlying the proposed construction.

In order to complete the proposed Bayesian formulation, we require priors for the model parameters in (1)–(3), namely σ_i^2 , for i = 1, ..., n, along with \mathbf{c}_{jt} and $\beta_{jt}^* = [\beta_{1jt}^*, ..., \beta_{K_{jt}jt}^*]^\mathsf{T}$, for each j = 1, ..., p, t = 1, ..., T. Regarding σ_i^2 , for i = 1, ..., n, we follow common practice and consider conditionally-conjugate Inv-Gamma (a_σ, b_σ) priors for each σ_i^2 , independently across i = 1, ..., n. Conversely, for \mathbf{c}_{jt} and β_{jt}^* we elicit priors that explicitly account for the temporal structure behind the evolution of both \mathbf{c}_{jt} and β_{jt}^* with periods t = 1, ..., T, for each j = 1, ..., p, thereby addressing *(ii)*. Focusing first on β_{jt}^* , this goal is accomplished by assuming independent Gaussian priors for the entries in such a vector, further centered around an higher-level mean function of time which is assigned a Gaussian process (GP) prior [see, e.g., Williams and Rasmussen, 2006]. More specifically, we let

$$(\boldsymbol{\beta}_{jt}^{\star} \mid \psi_{jt}, \delta_j^2) \sim \mathcal{N}_{K_{jt}}(\psi_{jt} \mathbb{1}_{K_{jt}}, \delta_j^2 \mathbf{I}_{K_{jt}}), \quad \text{independently for } j = 1, \dots, p, \ t = 1, \dots, T,$$

$$(\boldsymbol{\psi}_j = [\psi_{j1}, \dots, \psi_{jT}]^{\mathsf{T}} \mid \omega_j^2) \sim \mathcal{N}_T(\boldsymbol{\mu}_j, \omega_j^2 \boldsymbol{\Sigma}), \quad \text{independently for } j = 1, \dots, p,$$

$$(4)$$

where $\mathbb{1}_{K_{jt}}$ and $\mathbf{I}_{K_{jt}}$ denote the $K_{jt} \times 1$ vector of ones and the $K_{jt} \times K_{jt}$ identity matrix, respectively, whereas $\boldsymbol{\mu}_j$ and $\omega_j^2 \boldsymbol{\Sigma}$ are the mean vector and the covariance matrix induced by the GP prior on the finite time grid $t = 1, \ldots, T$. For $\boldsymbol{\Sigma}$ we consider, in particular, a squared exponential correlation function [see Williams and Rasmussen, 2006, Section 4] which allows the dependence between the generic ψ_{jt} and $\psi_{jt'}$ to progressively decrease as the distance between the time indexes t and t' increases. The mean vector $\boldsymbol{\mu}_j$ is instead elicited under a data-driven perspective (see Section 4), whereas δ_j^2 and ω_j^2 are assigned conditionally-conjugate Inv-Gamma (a_{δ}, b_{δ}) and Inv-Gamma (a_{ω}, b_{ω}) hyperpriors, respectively, independently for $j = 1, \ldots, p$.

As discussed above, temporal dependence is expected also in the sequence $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}$ of cluster assignment vectors, for each $j = 1, \ldots, p$. In particular, it is reasonable to assume that the generic $\mathbf{c}_{j,t-1}$ influences the formation of \mathbf{c}_{jt} through a mechanism in which only a subset of the countries change the corresponding cluster membership when moving from t - 1 to t, whereas the others preserve it. This Markovian dependence structure implies that $\operatorname{pr}(\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}) = \operatorname{pr}(\mathbf{c}_{j1})\operatorname{pr}(\mathbf{c}_{j2} \mid \mathbf{c}_{j1}) \cdots \operatorname{pr}(\mathbf{c}_{jT} \mid \mathbf{c}_{j,T-1})$, and hence, the mechanism defining $\operatorname{pr}(\mathbf{c}_{jt} \mid \mathbf{c}_{j,t-1})$ should be combined with a careful prior on the initial membership vector \mathbf{c}_{j1} at time t = 1, which induces a flexible characterization of the joint prior over the entire sequence $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}$, for each $j = 1, \ldots, p$. A construction of this type can be found in the temporal random partition (*t*RPM) prior recently proposed by Page et al. [2022]. Adapting this general construction to our specific context, we let

$$([\mathbf{c}_{j1},\ldots,\mathbf{c}_{jT}] \mid \alpha_j, M_j) \sim t_{\mathrm{RPM}}(\alpha_j, M_j), \qquad \text{independently for } j = 1,\ldots,p,$$
(5)

where $\alpha_i \in [0, 1]$ is a temporal persistence parameter controlling the transition mechanism behind $\operatorname{pr}(\mathbf{c}_{jt} \mid \mathbf{c}_{j,t-1})$, whereas $M_j \in \mathbb{R}_+$ regulates the formation of \mathbf{c}_{j1} through $\operatorname{pr}(\mathbf{c}_{j1})$, which in turn influences those of the subsequent vectors $\mathbf{c}_{j2}, \ldots, \mathbf{c}_{jT}$ under the Markovian structure discussed above. Specializing Page et al. [2022] to our setting, such a formation mechanism for c_{i1} is assumed to be driven by a Chinese restaurant process (CRP) prior with parameter M_j . This prior belongs to the general Gibbs-type class [e.g., De Blasi et al., 2015] and provides a flexible characterization for the formation of the grouping structures in \mathbf{c}_{j1} driven by a tractable urn scheme. In particular, let $\mathbf{c}_{j1}^{(-i)} = [c_{1j1}, \ldots, c_{i-1,j1}, c_{i+1,j1}, \ldots, c_{nj1}]$ be the cluster membership vector for the *j*-th spline basis at time t = 1, excluding the generic *i*-th country, and denote with $n_{ki1}^{(-i)}$ and $K_{i1}^{(-i)}$ the cardinality of cluster k and the total number of non-empty clusters in $\mathbf{c}_{i1}^{(-i)}$, respectively. Then, under this scheme, the prior on the cluster memberships for country i, given those of the others, coincides with $pr(c_{ij1} = k \mid \mathbf{c}_{j1}^{(-i)}) \propto n_{kj1}^{(-i)}$ if k is a cluster already occupied by the other n-1 countries, and $\operatorname{pr}(c_{ij1}=k \mid \mathbf{c}_{j1}^{(-i)}) \propto M_j$ if k is a new cluster (i.e., $k = K_{j1}^{(-i)} + 1$). Besides illustrating the tractability of this construction, along with the role of the parameter M_j , such an urn scheme also clarifies that the total number of cluster in \mathbf{c}_{i1} can be learned automatically, without the need to pre-specify it. This is a remarkable advantage compared to state-of-the-art cluster-based models for mortality [Hatzopoulos and Haberman, 2013, Léger and Mazzuco, 2021, Schnürch et al., 2021, Tang et al., 2022, Scognamiglio, 2022, Debon et al., 2023, Dimai, 2025] that cannot learn the number of clusters automatically, as part of the estimation process.

Under (5), the above prior for \mathbf{c}_{j1} is combined with a similarly-tractable assignment mechanism regulating the formation of $\mathbf{c}_{j2}, \mathbf{c}_{j3}, \ldots, \mathbf{c}_{jT}$ via the generic conditional law $\operatorname{pr}(\mathbf{c}_{jt} | \mathbf{c}_{j,t-1})$. This is accomplished by sampling a binary latent indicator variable $\gamma_{ijt} \sim \operatorname{Bern}(\alpha_j)$ that controls, for each country $i = 1, \ldots, n$, whether such country belongs to the same cluster when moving from $\mathbf{c}_{j,t-1}$ to \mathbf{c}_{jt} (i.e., $\gamma_{ijt} = 1$), or can possibly change its current allocation (i.e., $\gamma_{ijt} =$ 0). Given $\gamma_{jt} = [\gamma_{1jt}, \ldots, \gamma_{njt}]$ and $\mathbf{c}_{j,t-1}$, the allocation vector \mathbf{c}_{jt} at time t is sampled from the set of allocations compatible with $\mathbf{c}_{j,t-1}$, under γ_{jt} , namely all those allocation vectors yielding a partition of the n countries that can be obtained from the one encoded within $\mathbf{c}_{j,t-1}$ by reallocating only those countries for which $\gamma_{ijt} = 0$. In this sampling mechanism, each allocation vector within the compatible set is assigned a probability proportional to the one defined by the prior at time t = 1. Besides its simplicity, this construction provides an effective and flexible representation whose temporal persistence is directly regulated by α_j (the higher α_j is, the more similar $\mathbf{c}_{j,t-1}$ and \mathbf{c}_{jt} are). Inheriting the theory by Page et al. [2022] within our specific setting, this construction further ensures that the marginals for $\mathbf{c}_{j2}, \mathbf{c}_{j3}, \ldots, \mathbf{c}_{jT}$ have the same CRP prior as the one assumed above for \mathbf{c}_{j1} , thereby yielding a temporal construction which preserves, for all its marginals, all the advantages of the CRP formulation discussed above for \mathbf{c}_{j1} .

To conclude our Bayesian formulation, we select independent $\text{Beta}(a_{\alpha}, b_{\alpha})$ and $\text{Gamma}(a_M, b_M)$ hyperpriors for the quantities α_j and M_j in (5), respectively, for $j = 1, \ldots, p$. The $\text{Beta}(a_{\alpha}, b_{\alpha})$ hyperprior is considered also in Page et al. [2022], whereas the $\text{Gamma}(a_M, b_M)$ for the CRP concentration parameters, although not present in Page et al. [2022], is motivated by recent theoretical results in Ascolani et al. [2023] on the consistency properties of CRP constructions.

Section 3 clarifies that the above Bayesian formulation is also amenable to tractable posterior computation and inference leveraging a carefully-designed Gibbs-sampling algorithm.

3. Bayesian Computation and Inference

Posterior inference for the Bayesian model in Section 2 is conducted via Monte Carlo leveraging the draws of a carefully-designed Gibbs sampler targeting the posterior distribution of the model parameters given the observed mortality rates. Section 3.1 derives in detail such a Gibbs sampler, while Section 3.2 illustrates how the resulting posterior samples are leveraged to perform Monte Carlo inference on the local clustering structures among countries and the corresponding mortality patterns.

3.1 Gibbs sampler

The proposed Gibbs sampler iterates sequentially between two main steps. First, the temporal cluster allocations are sampled, along with the *t*RPM hyperparameters in (5), from the corresponding full-conditionals by adapting the algorithm of Page et al. [2022, see Section B, Supplementary Materials] to our B-splines construction, and the results in Escobar and West [1995] for the CRP hyperparameters M_j , $j = 1, \ldots, p$. Second, conditionally on the group allocations, the cluster-specific B-splines coefficients in (4) and the corresponding hyperparameters are updated leveraging Gaussian and inverse-Gamma conjugacy.

In order to update the temporal grouping structures in $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}$, for $j = 1, \ldots, p$, along with the parameters of the associated t_{RPM} prior, let us first recall that the generic allocation vectors $\mathbf{c}_{j,t-1}$ and \mathbf{c}_{jt} are compatible, with respect to γ_{jt} , if the partition of the *n* countries encoded within \mathbf{c}_{jt} may be derived from the one associated with $\mathbf{c}_{j,t-1}$ by reallocating only those countries for which $\gamma_{ijt} = 0$. Given $\mathbf{c}_{j,t-1}$ and γ_{jt} , define with $\mathbb{C}(\mathbf{c}_{j,t-1}, \gamma_{jt})$ the set comprising all partitions induced by \mathbf{c}_{jt} that are compatible with $\mathbf{c}_{j,t-1}$, under γ_{jt} , and let $\Gamma_{jt} = \{i = 1, \ldots, n : \gamma_{ijt} = 1\}$ be the set of countries whose group allocation does not change from t-1 to t. Furthermore, let us denote with $\mathbf{c}_{jt}^{\Gamma_{jt}}$ the partition induced by \mathbf{c}_{jt} , but considering only those countries with indexes in Γ_{jt} . Then, leveraging the results in Page et al. [2022], the full-conditional distribution for the latent indicators γ_{ijt} is a Bernoulli variable with probabilities

$$\operatorname{pr}(\gamma_{ijt} = 1 \mid -) = \frac{\alpha_j}{\alpha_j + (1 - \alpha_j)\operatorname{pr}(\operatorname{c}_{jt}^{\Gamma_{jt}^{(+i)}} \mid M_j)/\operatorname{pr}(\operatorname{c}_{jt}^{\Gamma_{jt}^{(-i)}} \mid M_j)} \mathbb{I}[\operatorname{c}_{j,t-1}^{\Gamma_{jt}^{(+i)}} = \operatorname{c}_{jt}^{\Gamma_{jt}^{(+i)}}], \tag{6}$$

independently for each i = 1, ..., n, j = 1, ..., p and t = 1, ..., T, where $\mathbb{I}[\cdot]$ denotes the indicator function, while $\Gamma_{jt}^{(-i)} = \Gamma_{jt} \setminus \{i\}$ and $\Gamma_{jt}^{(+i)} = \Gamma_{jt}^{(-i)} \cup \{i\}$. Note that in (6), the ratio $\operatorname{pr}(\mathbf{c}_{jt}^{\Gamma_{jt}^{(+i)}} \mid M_j)/\operatorname{pr}(\mathbf{c}_{jt}^{\Gamma_{jt}^{(-i)}} \mid M_j)$ can be computed in closed-form lever-

Note that in (6), the ratio $\operatorname{pr}(\mathbf{c}_{jt}^{\Gamma_{jt}^{i}} \mid M_j)/\operatorname{pr}(\mathbf{c}_{jt}^{\Gamma_{jt}^{i}} \mid M_j)$ can be computed in closed-form leveraging the results in Page et al. [2022] under the urn scheme of the CRP prior discussed in Section 2. This is also a key to update the allocations c_{ijt} of those countries for which $\gamma_{ijt} = 0$; if the sampled γ_{ijt} is 1, then c_{ijt} is kept fixed at the allocation drawn for *i* at the pair (j, t-1). To this end, let $\mathbf{c}_{jt}^{(-i)}$ be the vector of cluster allocations after removing the entry c_{ijt} , and denote with $\mathbf{c}_{jt}^{(c_{ijt}=k)}$ the membership vector $[c_{1jt}, \ldots, c_{ijt} = k, \ldots, c_{njt}]$. Furthermore, let $K_{jt}^{(-i)}$ be the total number of unique clusters in $\mathbf{c}_{jt}^{(-i)}$ and define with $r_{ixt}^{(j)} = \log m_{ixt} - \sum_{j'\neq j} \beta_{c_{ij't}j't}^{\star} g_{j'}(x)$ the partial residuals under model (1)–(3) without considering the *j*-th spline. Then, leveraging the Bayes rule and the CRP urn scheme, we have that the full-conditional distribution for those c_{ijt} having $\gamma_{ijt} = 0$ is a categorical variable with probabilities, for each $k = 1, \ldots, K_{jt}^{(-i)} + 1$, given by

$$\operatorname{pr}(c_{ijt} = k \mid -) \propto \\ \operatorname{pr}(c_{ijt} = k \mid \mathbf{c}_{jt}^{(-i)}, M_j) \mathbb{I}[\mathbb{c}_{j,t+1} \in \mathbb{C}(\mathbf{c}_{jt}^{(c_{ijt}=k)}, \gamma_{j,t+1})] \prod_{x \in \mathcal{X}} \phi(r_{ixt}^{(j)} - \beta_{kjt}^{\star} g_j(x); \sigma_i^2),$$

$$(7)$$

for every i = 1, ..., n, j = 1, ..., p and t = 1, ..., T, where $\phi(r_{ixt}^{(j)} - \beta_{kjt}^{\star}g_j(x); \sigma_i^2)$ is the density, evaluated at $r_{ixt}^{(j)} - \beta_{kjt}^{\star}g_j(x)$, of the zero-mean Gaussian distribution with variance σ_i^2 , whereas, as discussed in Section 2, $\operatorname{pr}(c_{ijt} = k \mid \mathbf{c}_{jt}^{(-i)}, M_j)$ can expressed under the CRP urn scheme as

$$\operatorname{pr}(c_{ijt} = k \mid \mathbf{c}_{jt}^{(-i)}, M_j) \propto \begin{cases} n_{kjt}^{(-i)} & k = 1, \dots, K_{jt}^{(-i)}, \\ M_j & k = K_{jt}^{(-i)} + 1, \end{cases}$$

where $n_{kjt}^{(-i)}$ denotes the size of the k-th cluster after removing unit i.

Given the samples of \mathbf{c}_{jt} and γ_{jt} , for all $t = 1, \ldots T$, the *t*RPM hyparameters α_j are updated from the full-conditional Beta distributions

$$(\alpha_j \mid -) \sim \text{Beta}\left(a_\alpha + \sum_{i=1}^n \sum_{t=1}^T \gamma_{ijt}, b_\alpha + nT - \sum_{i=1}^n \sum_{t=1}^T \gamma_{ijt}\right),\tag{8}$$

independently for j = 1, ..., p, whereas the CRP concentration parameters M_j , j = 1, ..., p, are updated following the data-augmentation scheme described in Escobar and West [1995].

Once the group membership vectors $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}$ have been updated for each $j = 1, \ldots, p$, it is possible to sample the cluster-specific coefficients in β_{jt}^{\star} , for $j = 1, \ldots, p$ and $t = 1, \ldots, T$, along with the corresponding time-varying means in ψ_j , for $j = 1, \ldots, p$. Combining prior (4) with the model (1)–(3), this can be accomplished by leveraging directly Gaussian-Gaussian conjugacy. In particular, the full-conditional distribution for the generic β_{kjt}^{\star} is

$$(\beta_{kjt}^{\star} \mid -) \sim \mathcal{N}(\omega_{\beta_{kjt}^{\star}}^{-1} \eta_{\beta_{kjt}^{\star}}, \omega_{\beta_{kjt}^{\star}}^{-1}), \tag{9}$$

independently across every $k = 1, \ldots, K_{jt}$, $j = 1, \ldots, p$ and $t = 1, \ldots, T$, where $\omega_{\beta_{kjt}^*} = 1/\delta_j^2 + \sum_{x \in \mathcal{X}} g_j^2(x) \sum_{i:c_{ijt}=k} (1/\sigma_i^2)$, while $\eta_{\beta_{kjt}^*} = (\psi_{jt}/\delta_j^2) + \sum_{i:c_{ijt}=k} (1/\sigma_i^2) \sum_{x \in \mathcal{X}} r_{ixt}^{(j)} g_j(x)$. Similarly, the full-conditional for each vector ψ_j is

$$(\boldsymbol{\psi}_j \mid -) \sim \mathrm{N}_T(\boldsymbol{\Omega}_{\psi_j}^{-1} \boldsymbol{\eta}_{\psi_j}, \boldsymbol{\Omega}_{\psi_j}^{-1}), \tag{10}$$

for j = 1, ..., p, with $\mathbf{\Omega}_{\psi_j} = \omega_j^{-2} \mathbf{\Sigma}^{-1} + \delta_j^{-2} \operatorname{diag}(K_{j1}, ..., K_{jT})$ and $\mathbf{\eta}_{\psi_j} = \omega_j^{-2} \mathbf{\Sigma}^{-1} \boldsymbol{\mu}_j + \delta_j^{-2} \bar{\boldsymbol{\beta}}_j$, where $\bar{\boldsymbol{\beta}}_j$ is a vector of dimension $T \times 1$ having generic *t*-th entry $\sum_{k=1}^{K_{jt}} \beta_{kjt}^{\star}$.

To conclude the Gibbs-sampling routine it remains to update the variance parameters σ_i^2 , $i = 1, \ldots, n$ in (1), along with δ_j^2 and ω_j^2 , $j = 1, \ldots, p$, entering the priors in (4). By conditioning on the quantities sampled in (7) and (9)–(10), the updates for these variance parameters follow directly from inverse-Gamma conjugacy properties, thereby obtaining

$$(\sigma_i^2 \mid -) \sim \text{Inv-Gamma}\left(a_\sigma + XT/2, b_\sigma + (1/2)\sum_{x \in \mathcal{X}} \sum_{t=1}^T [\log m_{ixt} - f_{it}(x)]^2\right).$$
 (11)

for every i = 1, ..., n, with $f_{it}(x)$ defined as in (2)–(3), and

$$(\delta_{j}^{2} \mid -) \sim \text{Inv-Gamma} \left(a_{\delta} + \sum_{t=1}^{T} K_{jt}/2, b_{\delta} + (1/2) \sum_{t=1}^{T} \sum_{k=1}^{K_{jt}} (\beta_{kjt}^{\star} - \psi_{jt})^{2} \right),$$

$$(\omega_{j}^{2} \mid -) \sim \text{Inv-Gamma} \left(a_{\omega} + T/2, b_{\omega} + (\psi_{j} - \mu_{j})^{\top} \Sigma^{-1} (\psi_{j} - \mu_{j})/2 \right),$$

$$(12)$$

for every $j = 1, \ldots, p$.

3.2 Monte Carlo inference

Leveraging the samples produced by the Gibbs routine outlined in Section 3.1 posterior inference on the quantities of interest proceeds via Monte Carlo. As discussed in Sections 1–2, within our context a primary focus is on inferring grouping structures among countries induced by similarities among the corresponding mortality rates, and how these group structures vary locally across ages and periods. This information is contained in the posterior samples of the allocation vectors \mathbf{c}_{jt} for $j = 1, \ldots, p$ and $t = 1, \ldots, T$, which we summarize through the $n \times n$ posterior co-clustering matrices $\hat{\mathbf{P}}_{jt}$, $j = 1, \ldots, p$, $t = 1, \ldots, T$, whose generic element $\hat{\mathbf{P}}_{jt[i,i']}$ is defined as the relative proportion of Gibbs samples in which countries i and i' have the same group allocation, at the pair (j, t). This provides an estimate of the posterior probabilities of co-clustering that is useful for quantifying uncertainty in \mathbf{c}_{jt} for $j = 1, \ldots, p$, $t = 1, \ldots, T$, beyond single point estimates. As such, these matrices will be object of study in the mortality data application in Section 5.

When a single point estimate $\hat{\mathbf{c}}_{jt}$ of \mathbf{c}_{jt} is of interest for each $j = 1, \ldots, p$ and $t = 1, \ldots, T$, the above *posterior co-clustering matrices* can be summarized under the decision-theoretic framework

of Wade and Ghahramani [2018] to obtain the estimate

$$\hat{\mathbf{c}}_{jt} = \operatorname{argmin}_{\mathbf{c}'_{jt}} \mathbb{E}_{\mathbf{c}_{jt}} [\operatorname{VI}(\mathbf{c}_{jt}, \mathbf{c}'_{jt}) \mid \log m], \quad j = 1, \dots, p, \ t = 1, \dots, T,$$

where log m is the array of log-mortality rates observed for the n countries across all ages and periods, whereas VI is the variation of information distance [Meilă, 2007], namely a metric measuring the dissimilarity among two generic allocation vectors \mathbf{c}_{jt} and \mathbf{c}'_{jt} based on the associated individual and joint entropies. In practice, the above minimization problem is solved, for each j = $1, \ldots, p$ and $t = 1, \ldots, T$, via R package mcclust.ext [Wade and Ghahramani, 2018] leveraging, as input, the corresponding *posterior co-clustering matrices* $\hat{\mathbf{P}}_{jt}, j = 1, \ldots, p, t = 1, \ldots, T$.

Besides the grouping structures encoded in $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{jT}$, for $j = 1, \ldots, p$, it is also of interest to study the trajectories $\beta_{ij1}, \ldots, \beta_{ijT}$ of the B-splines coefficients. Recalling (2), these trajectories characterize the temporal evolution of the mortality levels for country *i* at the age interval associated with the *j*-th spline. Posterior samples for these coefficients can be derived directly from those for \mathbf{c}_{jt} and β_{jt}^* , after noticing that, under (3), $\beta_{ijt} = \beta_{c_{ijt}jt}^*$, for every $i = 1, \ldots, n, j =$ $1, \ldots, p$ and $t = 1, \ldots, T$. As such, point estimates for the dynamic country-specific B-spline coefficients can be derived via Monte Carlo by computing the average of the resulting samples for each $\beta_{ijt}, i = 1, \ldots, n, j = 1, \ldots, p$ and $t = 1, \ldots, T$.

4. Simulation Study

Before employing the proposed model in the motivating mortality data application, we first assess its performance in recovering the *true* data-generative structures in a simulation study. Recalling Sections 1–3, the overarching focus is on quantifying to what extent the proposed model can learn accurately realistic grouping structures that vary locally, along with the associated cluster-specific parameters, in different combinations of ages and periods.

Consistent with the above goal, we simulate synthetic log-mortality rates, under the model outlined in (1)–(3) focusing on n = 5 countries across T = 10 periods and for ages $\mathcal{X} = \{0, 1, \ldots, 100\}$. More specifically, $\log m_{ixt}$ are simulated as in (1) with $\sigma_i = \sigma = 0.05$, for $i = 1, \ldots, 5$, and $f_{it}(x)$ defined through (2)–(3) considering p = 6 quadratic B-splines whose cluster-specific coefficients are generated via (4), setting $\delta_j = 0.05$ for $j \neq 5$, $\delta_5 = 0.1$, and letting ψ_j , $j = 1, \ldots, p$ correspond to parallel decreasing lines with slope -0.02. To assess the ability of the proposed model in inferring complex local group structures among countries, we do not produce $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{j10}$, for $j = 1, \ldots, 6$ from the assumed prior. Rather, we set the group allocations manually as in Figure 2, in order to explore a wide spectrum of time-varying local clustering patterns. For example, for the age classes associated with the spline bases 3 (top-right panel) and 4 (bottom-left panel) the synthetic countries are grouped into stable clusters across time, albeit with differing co-clustering patterns. Conversely the remaining spline bases exhibit more complex dynamic clustering patterns (see, e.g., spline basis 6 where the synthetic countries often change cluster membership).

Leveraging the above simulated data, we perform Bayesian inference under the model proposed in Section 2, setting diffuse hyperparameters $a_{\sigma} = b_{\sigma} = 10^{-3}$, $a_{\delta} = b_{\delta} = a_{\omega} = b_{\omega} = 10^{-3}$, $a_M = 2 \cdot 10^{-3}$, $b_M = 10^{-3}$ and $a_{\alpha} = 1$, $b_{\alpha} = 1$. Consistent with standard practice in Gaussian processes



Figure 2: Cluster assignments in the simulation study. Colors and numbers represent true cluster memberships.

literature [see, e.g., Williams and Rasmussen, 2006], the entries of Σ in (4) are defined through a squared-exponential kernel with length scale 1.5, i.e. $\Sigma_{[t,t']} = \exp[-0.5(t-t')^2/(1.5)^2]$, for every (t,t'). Finally, to achieve an improved calibration of the proposed model, the mean vectors $\mu_j = [\mu_{j1}, \ldots, \mu_{j10}]^{\mathsf{T}}$, $j = 1, \ldots, 6$, in (4) are defined in a data-driven manner. This is accomplished by first obtaining, separately for every $t = 1, \ldots, 10$, an OLS estimate of the splines coefficients under model (1) applied to data log m_{ixt} , $i = 1, \ldots, 5$, $\mathcal{X} = \{0, 1, \ldots, 100\}$, with $f_{ix}(t) = \sum_{j=1}^{6} \beta_{jt}g_j(x)$. For each $j = 1, \ldots, 6$, the resulting estimates $\hat{\beta}_{j1}, \ldots, \hat{\beta}_{j10}$ are subsequently smoothed via LOESS to obtain the desired data-driven specification for the entries $\mu_{j1}, \ldots, \mu_{j10}$ of μ_j . All these settings and hyperparameter choices proved robust also the in the mortality data application in Section 5, and moderate changes in such quantities did not substantially affect the final conclusions and the performance in the simulation.

Under the above settings, posterior inference for the proposed model proceeds via Monte Carlo as outlined in Section 3.2, leveraging the samples produced by the Gibbs routine derived within Section 3.1. Such a routine is run for 20,000 iterations, discarding the first 10,000 as a conservative burn-in. Traceplots and autocorrelation plots indicate satisfactory mixing of the chains. In fact, in this simulation study we observe convergence much before the burn-in employed. Nonetheless, we opted for a more conservative setting that can be considered as a default in general contexts, including in the motivating mortality data application in Section 5.

	t = 1	t=2	t = 3	t = 4	t = 5	t = 6	t = 7	t = 8	t = 9	t = 10
Spline 1 $(j = 1)$	0.991	0.998	0.999	0.995	0.976	1.000	1.000	1.000	1.000	1.000
Spline 2 $(j = 2)$	0.995	0.999	0.999	1.000	0.998	0.991	0.999	1.000	1.000	0.995
Spline 3 $(j = 3)$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spline 4 $(j = 4)$	0.991	1.000	0.999	1.000	0.997	1.000	1.000	0.999	0.987	0.990
Spline 5 $(j = 5)$	0.996	1.000	1.000	1.000	0.994	0.997	1.000	0.999	0.999	0.993
Spline 6 $(j = 6)$	0.914	0.897	0.968	0.969	0.982	0.979	0.983	0.983	0.989	0.990

Table 1. Simulation study. Posterior means of co-clustering accuracies for each combination (j, t).



Figure 3: Simulation study. Posterior means of $\beta_{ij1}, \ldots, \beta_{ij10}$ (lines) and true values (points) for each country $i = 1, \ldots, 5$ (panels) and basis function $g_j(x), j = 1, \ldots, 6$ (colors).

Consistent with our overarching focus, we first assess in Table 1 to what extent the proposed model is able to learn the *true* clustering structures displayed in Figure 2 among the 5 synthetic countries. To provide a comprehensive assessment of the clustering accuracy achieved by the posterior for $\mathbf{c}_{j1}, \ldots, \mathbf{c}_{j10}, j = 1, \ldots, 6$, beyond the one obtained under a single point estimate, we compute, for each Gibbs sample of $\mathbf{c}_{jt}, t = 1, \ldots, 10, j = 1, \ldots, 6$, the percentage of pairs of countries that are correctly co-clustered. The posterior means of these percentages over the 10,000 retained samples is reported in Table 1, and confirm the excellent performance of the proposed model in recovering the co-clustering patterns induced by the *true* group structures within Figure 2. In particular, the accuracy measures are above 0.975 for all spline bases and time points, except for 6-th one where we observe a slight performance deterioration. This result is expected since such a basis presents the most complex clustering patterns, with frequent changes of group memberships for all countries. Nonetheless, even in this highly challenging regime, we still observe remarkable performance with all accuracy measures across time points above 0.89.

As shown within Figure 3, the above remarkable performance in learning local clustering structures directly translates into highly accurate point estimates for the country-specific coefficient trajectories $\beta_{ij1} = \beta^*_{c_{ij1}j1}, \ldots, \beta_{ij10} = \beta^*_{c_{ij10}j10}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 6$. These point estimates are obtained as detailed in Section 3.2, and compared with the associated *true* values within Figure 3. Results confirm the ability of the model to characterize the underlying trajectories of each spline coefficient accurately, even for complex underlying temporal dynamics. Under (1)–(3), this implies effective learning of the data-generative mechanism for the synthetic log-mortality rates.

5. Local Clustering of Age-Period Mortality Surfaces for 14 Countries

We conclude by showcasing the performance of the proposed model in learning local clustering structures across ages and periods induced by the log-mortality rates of 14 countries (Australia, Belgium, Canada, Switzerland, Denmark, Spain, Finland, France, Italy, the Netherlands, Norway,



Figure 4: Graphical representation of the selected B-splines bases $g_1(x), \ldots, g_{20}(x)$. The number associated to each spline $g_j(x)$ denotes the age at which such a spline takes its maximum value.

Sweden, the United Kingdom and the United States). The original data $\log m_{ixt}$ are retrieved from the Human Mortality Database [Human Mortality Database, 2024] for ages from 0 until 98 years-old, over a time horizon of 88 years (1933 – 2020).

In applying the proposed model to the above data we follow standard practice in multi-country studies [e.g., Li and Lee, 2005, Aliverti et al., 2022] by considering separate analyses for male and female sub-populations. Posterior inference proceeds under the same hyperparameters and Gibbs settings as those considered for the simulation studies in Section 4, except for the choice of the B-spline bases that are set as in Figure 4 to achieve a more realistic characterization of the observed age patterns of mortality. This choice allows increased flexibility in those age ranges where larger local variations are expected (i.e., infant and senescent groups), and has proved effective also in recent single-country analyses [e.g., Pavone et al., 2024], thereby motivating its use also in the newly-developed multi-country model. As in the simulations studies, the traceplots for the quantities of interest and the convergence diagnostics did not provide evidence against convergence. Section 5.1 summarizes the results of posterior inference under the proposed model, whereas Section 5.2 highlights previously-unexplored local clustering structures with a specific focus on the United States. Finally, Section 5.3 explores possible associations between the novel co-clustering patterns inferred by the proposed model and relevant socio-economic variables, including the gross domestic product (GDP) and health expenses.

5.1 Results

Figure 5 shows the evolution of posterior means for the number of clusters inferred by the proposed model across the age intervals associated with the 20 spline bases within Figure 4, over the temporal window analyzed. Such a quantity is generally stable for both infant mortality (age 0) and adult/elder mortality (from age 60 to 90). Instead, a considerable variability in the number of clusters is observed for children and adolescents (from age 1 to 25) and late mortality (from age 91), with different patterns. Besides supporting the need of allowing the group structures among countries to vary locally with ages and periods, these trends can be interpreted as a measure of variability in mortality rates across countries, with larger number of groups corresponding to age intervals where the different countries display higher dissimilarities over periods. Consistent with this interpretation, the results in Figure 5 suggest that senescent mortality is characterized by an



Figure 5: Time trajectories for the posterior mean of the number of clusters at the age intervals associated with the 20 selected B-spline bases.

increasing level of similarity among countries over the recent years, whereas child mortality exhibits an opposite trend after the early 1960s, indicating a divergence among countries that might have been driven by social phenomena such as the baby boom. Overall, males and females are characterized by similar patterns, with few notable exceptions. The number of clusters for elders (ages 91 - 98) converges to few groups for both sub-populations, although males are generally characterized by higher dissimilarities among countries since the first period under investigation. It is also interesting to notice how the mortality of young and adult males (age 16 - 40) is affected by World War II during the 1940s, and reaches a plateau only after the peak associated with the military services.

The ability of the proposed model to effectively capture local clustering patterns is illustrated in Figure 6, where the observed infant log-mortality rates in Italy, Sweden, the UK and the USA are compared with the corresponding probabilities of co-clustering estimated from the samples of the Gibbs algorithm presented in Section 3.1. This specialized analysis is useful to further illustrate the main advantages of the newly-proposed approach and its implications for the global characterization of mortality rates. In particular, the right panel of Figure 6 indicates a fluctuating probability of co-clustering between the UK and the USA (bottom row), with larger values before 1940, in the early 1950s, and more recently from 1970 to 1980, in agreement with the observed trends of mortality rates reported in the left panel. Evidence of co-clustering is observed also for Italy with both the UK and the USA around the 1980s. In contrast, Sweden is characterized by a peculiar and separate trajectory, that overlaps with Italy only in recent years (top row).

To fully explore the evolution of the local co-clustering patterns, beyond the above specialized analysis, Figures 7a and 7b display the estimated co-clustering probabilities across all pairs of countries, for males and females respectively. Both figures illustrate the co-clustering probabilities



Figure 6: Observed male infant log-mortality rates for Italy (ITA), Sweden (SWE), the United Kingdom (UK) and the United States (USA) (left), and estimated dynamic co-clustering probabilities for the corresponding pairs (right).

for every B-spline basis (corresponding to a different age interval) through a matrix with rows denoting all pairs of countries and columns referring to calendar years. Colors range from blue to red as the estimated co-clustering probabilities vary from 0 to 1. The results in Figures 7a and 7b indicate, in general, an increasing overlap among countries over periods in terms of the associated age-specific log-mortality rates. Interestingly, evidence of increasing co-clustering is progressively more present both as time advances and as the population ages, demonstrating a dual-directional reinforcement of these similarities. For instance, in the eldest age interval (96–98) of the male subpopulation, the proportion of country pairs with a high probability of co-clustering has steadily risen in recent years, forming a distinct block of countries with large probabilities of co-clustering in the bottom-right panel of Figure 7a. Such compression is stronger for older ages, with young age classes demonstrating smaller probabilities of co-clustering than elder ones. Furthermore, this phenomenon finds empirical evidence in both the male (Figure 7a) and female (Figure 7b) subpopulations, thereby supporting previous findings on demographic convergence [e.g., Vaupel et al., 2011, Wilson, 2011].

Focusing on child mortality, the younger age classes (from age 1 to 19) showcase a neat temporal pattern, with most of the co-clustering concentrated between the late 1950s and the early 1980s. This behavior is coherent with the evolution of the number of clusters in Figure 5, and suggests that the countries under investigation have been characterized by different levels of child mortality until World War II, followed by the rapid improvement of mortality rates in the early 1960s that created few common clusters with similar mortality patterns. After the 1980s, the co-clustering structures become more irregular and countries deviate to more individual trends. This finding is worth future investigations.

The co-clustering patterns among countries for adult mortality are reported within the third row of Figure 7a and Figure 7b, and generally show a common diagonal structure with some groups

5.1 5.1 Results





(b) Female population

Figure 7: Evolution of the estimated co-clustering probabilities for country pairs (rows) and years (columns), over all age intervals (panels). Colors correspond to the estimated probabilities, and range from blue (low) to red (high).

of European countries experiencing important changes of cluster membership around the 1980s. This phenomenon can be more clearly appreciated, for the male sub-population, in the left panel of Figure 8, that reports the estimates of the dynamic spline coefficient associated with the ages



Figure 8: Estimates of β_{14} and β_{19} (age classes 61–79 and 91–97) in the male population.

61-79. Such estimates show how the Netherlands and Denmark are characterized by low mortality rates until the 1970s, while Belgium, Finland and the UK experienced some of the highest rates during that period. Later on, mortality drops rapidly for most countries, whereas the Netherlands and Denmark experience increments that makes these countries co-cluster with higher mortality ones. Structural changes in cross-country relations characterize this period. The potential impact of such changes is visible in the right panel of Figure 8, which displays the evolution of the spline coefficient associated with the age interval 91-97, and illustrates that after 1970s mortality levels of elder ages collapse into fewer clusters with stable trends for several years.

To further study the similarity between the group structures among countries inferred for the male and female log-mortality rates, we have also computed the normalized variation of information distance (NVI) [e.g., Wade and Ghahramani, 2018] among the group membership vectors estimated for these two sub-populations. Overall, the similarity between the partitions is more evident for the two youngest age classes (0 and 1-4), with values of the NVI below 0.44 and 0.58, respectively, throughout the entire time window. For example, female infant mortality in 2020 is divided into four groups. The first comprises the USA, the second Canada, the third Central European countries (Belgium, Switzerland, Denmark, France, the Netherlands and the UK) and Australia, and the fourth encodes both Mediterranean Europe (Italy, Spain) and Scandinavian countries (Finland, Norway, Sweden). The corresponding partition for the male sub-population consists, instead, of five groups, with the same first three as for the female sub-population, and the remaining two obtained by splitting Mediterranean Europe and Scandinavia into two separate clusters. More remarkable differences between male and female sub-populations are observed among middle-aged individuals, with NVI values above 0.5 after the 1960s. Interestingly, for the elderly population, a trend of increasing similarity started in the late 1980s, mirroring the patterns observed for young ages. Overall, these trends reflect a progressive convergence of mortality patterns across countries [Vaupel et al., 2011] for both males and females.

5.2 A specialized focus on the United States

We devote here special attention to the analysis of the co-clustering probabilities between the USA and the other countries, estimated under the proposed model. This specialized focus is motivated by the fact that the USA has exhibited peculiar patterns in recent years, with rising mortality rates among certain demographic groups and a decline of life expectancy [see, e.g., Bergeron-Boucher et al., 2020, Case and Deaton, 2021, Glei, 2022]. This phenomenon has been extensively studied in the literature, with a general consensus attributing these unexpected mortality increments to factors such as persistent disparities in healthcare access, increasing suicide rates, and more recently, to the opioid epidemic [see, e.g., Woolf and Schoomaker, 2019]. To complement and extend these findings, it is therefore particularly interesting to study how the USA co-clusters with the other countries under investigation in terms of mortality rates.

Consistent with the above goal, Figure 9a displays the dynamic probabilities of co-clustering among the USA and the other countries at selected age intervals for the male sub-population, estimated under the proposed model. The results in Figure 9a point toward a persistent co-clustering among the USA and Finland. This finding indicates strong similarities in terms of premature male mortality and can be attributed to suicide incidence and cardiovascular diseases. Indeed, although Scandinavian countries provide universal and publicly funded welfare systems for all citizens, significant disparities still persist across the socio-economic spectrum due to the so-called "Nordic Paradox" [see, e.g., Mackenbach, 2017, Højstrup et al., 2023]. As such, Figure 9a suggests that the "Death of Despair" phenomenon in the USA [e.g., Case and Deaton, 2021, Glei, 2022] might have interesting similarities with the "Nordic Paradox".



Figure 9: Estimated dynamic co-clustering probabilities between the USA and the other countries, for selected age intervals. Colors range from light to dark as the probability varies from low to high.

Focusing on the female sub-population, Figure 9b provides evidence of a relatively-persistent co-clustering among the USA and Denmark females at adult ages, varying across time and with age intervals. Such a peculiar result should be further investigated, since the interwar generation of Danish females represents a relevant demographic group that comprises cohorts who experienced a peculiar stagnation in life expectancy, resulting in a divergent trajectory compared to other Scand-inavian countries [Lindahl-Jacobsen et al., 2016]. As such, recent improvements in life expectancy can be linked to a cohort effect associated with this group that is visible also from the co-clustering patterns in Figure 9b. Interestingly, Belgian females join these clusters in specific age groups and time periods; a similar trend can be observed for UK at age 41-69. Such patterns might be related to higher proportion of diseases of the circulatory system for Belgian females [Bergeron-Boucher et al., 2020] and to general declines in life expectancy in the UK [Ho and Hendi, 2018].

5.3 Associations among local clustering structures and socio-economic indicators

The results in Sections 5.1–5.2 showcase relevant co-clustering structures among countries. These structures display convergence phenomena over periods for specific age classes [e.g., Oeppen and Vaupel, 2002], along with evidence of growing disparities for other ages. Available studies suggest that such patterns are driven by socio-economic factors such as quality of healthcare, education, and life standards [Marmot, 2005, Vallin and Meslé, 2004, European Commission and Executive Agency for Health and Consumers, 2013].

To provide additional empirical support to the above studies, we conclude our analysis by assessing to what extent the group structures inferred by the proposed model are associated with relevant socio-economic indicators, covering, in particular, the per-capita gross domestic product (GDP) in USD, unemployment rate, health expenditure (as the percentage of a country GDP), nutrition and child vaccination rates (against diphtheria, tetanus, pertussis); refer to OECD [2024] for more details. Data on these indicators are available, for all countries, across varying time periods. The earliest data for the GDP date back to 1970, whereas child vaccination rates and health spending are available starting from 1990. Finally, data on nutrition quality and unemployment rate start from 2010. Following standard practice, the association between these socio-economic indicators and the clustering structures among countries inferred by the proposed model is evaluated using the $\eta^2 \in [0,1]$ coefficient, which quantifies the variability of each indicator between the inferred clusters with respect to the total variability of such an indicator. Values of the η^2 close to 1 imply that the group structures among countries learned on the basis of similarities in the associated mortality rates are also explicative of the variability for the indicator under analysis, thereby suggesting possible associations. Notice that this measure does not inform on the direction of such an association, nor on possible causal interpretations. Nonetheless, it offers a sensible perspective for prioritizing specific indicators within explanatory studies on the socio-economic determinants of multi-country mortality patterns.

Figure 10 reports, for both the female and male sub-population, the posterior means of the η^2 coefficients in matrix form. Rows correspond to the age intervals $j = 1, \ldots, p$ associated with the different spline bases, columns to calendar years $t = 1, \ldots, T$, and panels to socio-economic indicators $q = 1, \ldots, Q$. To obtain these posterior means we first compute the η^2 between each posterior



Figure 10: Posterior mean of η^2 coefficients between selected socio-economic indicators and the inferred mortalitybased clusterings in the female and male sub-populations.

sample of $\mathbf{c}_{jt} = [c_{1jt}, \ldots, c_{njt}]$ and the vector $\mathbf{w}_t^{(q)} = [w_{1t}^{(q)}, \ldots, w_{nt}^{(q)}]$, with generic element $w_{it}^{(q)}$ denoting the value of the socio-economic indicator q, for country i in period t. This produces posterior samples of the η^2 for each age interval j = 1, ..., p, period t = 1, ..., T and socio-economic indicator $q = 1, \ldots, Q$. Averaging over these samples yields the posterior means in the matrices in Figure 10, which highlight a general concordance in the η^2 coefficients for the female and male sub-populations with non-trivial associations between the inferred clusters and the socio-economic indicators analyzed, particularly at young ages, and with different patterns across calendar years. For example, nutrition, GDP, health spending and unemployment rate appear to be associated with mortality clusters through a lower-diagonal cohort structure spanning a large spectrum of age classes, from childhood until young-adults and adults. This suggests that past incentives have had a lasting impact on specific cohorts, ultimately improving the lifespan of individuals who directly benefited from these initiatives. Not surprisingly, child vaccination rates and nutrition are generally less associated with clusters observed at age 0. In fact, such interventions cannot prevent neonatal and infant mortality, which are caused by specific factors [such as, birth defects, sudden infant death syndrome or accidents; see MacDorman et al., 2013, Eberstein et al., 1990]. Instead, these indicators show a stronger association with the inferred clusters from age 1 until late adolescence, consistent with the protective effects of vaccination and proper nutrition in such ages.

6. Conclusions and Future Research Directions

Available statistical models for mortality data are either designed for analyzing single countries in isolation or for inferring global group structures among multiple countries with respect to the entire age-period mortality surface. This perspective prevents from unveiling more nuanced similarity patterns that are observed in practice over specific combinations of ages and calendar years, thereby limiting the possibility to learn and quantify relevant demographic phenomena localized at specific age classes and periods.

In this article, we overcome the above limitations through a novel multi-country model that characterizes the age pattern of mortality via a flexible B-spline expansion, and incorporates both temporal and age-specific clustering structures by allowing the coefficients of the B-spline bases to change in time via separate temporal random partition priors with cluster-specific Gaussian processes regulating the dynamic evolution of such coefficients. The resulting Bayesian formulation is amenable to tractable posterior inference via a Gibbs-sampling algorithm that facilitates interpretable reconstruction of group structures among countries, varying locally with both ages and periods. The unique advantages of the newly-proposed model are illustrated in simulation studies and in an application to mortality data for 14 OECD countries, where our formulation unveils fundamental local clustering structures, including unexplored ones that motivate future research in the area. For example, the in-depth analysis of the USA clustering patterns in Section 5.2 reveals peculiar similarities with specific Scandinavian countries for adult populations. In particular, the persistent similarity with Danish female log-mortality rates motivates further investigation, as this sub-population is known to have exhibited peculiar mortality behavior due to significant smoking prevalence during World War II. Furthermore, the inferred associations between the local clustering patterns reconstructed by the proposed model and relevant socio-economic indicators (see Section 5.3), highlight peculiar cohort effects that are worth future analyses.

Future research includes applying the proposed model to a broader range of countries, while refining the analysis on the association among the groups structures induced by mortality patterns and central socio-economic indicators. Both perspectives require overcoming challenges related to the availability of historical data. In fact, the 14 countries considered in this article are the only ones providing mortality rates that date back to our study period; analyzing more countries would imply constraining the analysis to narrower time windows. Improving inference on the associations with socio-economic indicators would require, instead, including such information directly within the proposed model. A possibile direction for addressing this goal is to combine the tRPM prior with a reinforcement mechanism favoring the formation of local clusters that are homogeneous also with respect to the associated socio-economic indicators. This could be accomplished by extending the combination among product partition models with covariates [Müller et al., 2011] and tRPM priors proposed in Page et al. [2022] to the case of dynamic covariates.

Finally, let us emphasize that although our model is motivated by demographic applications, the constructions and results in Sections 2–3 have broader methodological impact, and can be applied whenever interest lies in the detection of localized overlaps among surfaces associated with different populations. To our knowledge, methodological results in these directions are limited.

Acknowledgments

This research is supported by the MUR-PRIN 2022 project "CARONTE" (Prot. 2022KBTEBN), funded by the European Union – Next Generation EU, Mission 4, CUP: J53D23009400001. The original data analyzed are available at https://www.mortality.org/. The authors would also like to thank Stefano Mazzuco for the insightful discussion on an early version of this manuscript.

References

- A. Alexopoulos, P. Dellaportas, and J. J. Forster. Bayesian forecasting of mortality rates by using latent Gaussian models. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(2):689–711, 2019.
- E. Aliverti, S. Mazzuco, and B. Scarpa. Dynamic modelling of mortality via mixtures of skewed distribution functions. Journal of the Royal Statistical Society: Series A (Statistics in Society), 185(3):1030-1048, 2022.
- K. Antonio, A. Bardoutsos, and W. Ouburg. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, 5:245–281, 2015.
- F. Ascolani, A. Lijoi, G. Rebaudo, and G. Zanella. Clustering consistency with Dirichlet process mixtures. *Biometrika*, 110(2):551–558, 2023.
- M.-P. Bergeron-Boucher, J. M. Aburto, and A. van Raalte. Diversification in causes of death in low-mortality countries: emerging patterns and implications. *BMJ Global Health*, 5(7):e002414, 2020.
- C. G. Camarda. Smooth constrained mortality forecasting. Demographic Research, 41:1091-1130, 2019.
- A. Case and A. Deaton. Deaths of Despair and the Future of Capitalism. Princeton University Press, 2021.
- I. D. Currie. On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, 2016(4):356–383, 2016.
- I. D. Currie, M. Durban, and P. H. Eilers. Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4): 279–298, 2004.
- P. De Blasi, S. Favaro, A. Lijoi, R. H. Mena, I. Prünster, and M. Ruggiero. Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37: 212–229, 2015.
- P. De Jong and L. Tickle. Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, 13(1): 1–18, 2006.
- A. Debon, S. Haberman, and G. Piscopo. Multipopulation mortality analysis: bringing out the unobservable with latent clustering. Quality & Quantity, 58:1–17, 2023.
- M. Dimai. Clustering of mortality paths with the hellinger distance and visualization through the distatis technique. Statistical Methods & Applications, (In press):1–40, 2025.
- Y. Dong, F. Huang, H. Yu, and S. Haberman. Multi-population mortality forecasting using tensor decomposition. Scandinavian Actuarial Journal, 2020(8):754–775, 2020.
- I. W. Eberstein, C. B. Nam, and R. A. Hummer. Infant mortality by cause of death: main and interaction effects. Demography, 27:413-430, 1990.
- P. H. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. Statistical Science, 11(2):89–121, 1996.
- V. Enchev, T. Kleinow, and A. J. Cairns. Multi-population mortality models: fitting, forecasting and comparisons. Scandinavian Actuarial Journal, 2017(4):319–342, 2017.
- M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association, 90(430):577–588, 1995.
- European Commission and Executive Agency for Health and Consumers. Health Inequalities in the EU Final Report of a Consortium Consortium Lead: Sir Michael Marmot. European Commission, 2013.
- D. A. Glei. The US midlife mortality crisis continues: excess cause-specific mortality during 2020. American Journal of Epidemiology, 191(10):1677–1686, 2022.
- S. Haberman and A. Renshaw. A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics, 48(1):35–55, 2011.
- P. Hatzopoulos and S. Haberman. Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, 52(2):320–337, 2013.
- J. Y. Ho and A. S. Hendi. Recent trends in life expectancy across high income countries: retrospective observational study. BMJ, 362, 2018.
- S. Højstrup, J. H. Thomsen, and E. Prescott. Disparities in cardiovascular disease and treatment in the nordic countries. *The Lancet Regional Health-Europe*, 33, 2023.
- Human Mortality Database. University of California Berkeley, USA and Max Planck Institute for Demographic Research, Germany. 2024. URL http://www.mortality.org/.
- A. Hunt and D. Blake. On the structure and classification of mortality models. North American Actuarial Journal, 25(sup1):S215–S234, 2021.
- R. J. Hyndman and M. S. Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis, 51(10):4942–4956, 2007.
- R. J. Hyndman, H. Booth, and F. Yasmeen. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50:261–283, 2013.
- T. Kleinow. A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147–152, 2015.

- K. K. Lam and B. Wang. Multipopulation mortality modelling and forecasting: the weighted multivariate functional principal component approaches. *Journal of Applied Statistics*, 50(15):3177–3198, 2023.
- R. Lee and T. Miller. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549, 2001.
- R. D. Lee and L. R. Carter. Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671, 1992.
- A.-E. Léger and S. Mazzuco. What can we learn from the functional clustering of mortality data? An application to the Human Mortality Database. *European Journal of Population*, 37:769–798, 2021.
- N. Li and R. Lee. Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. Demography, 42:575–594, 2005.
- N. Li, R. Lee, and P. Gerland. Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051, 2013.
- R. Lindahl-Jacobsen, R. Rau, B. Jeune, V. Canudas-Romo, A. Lenart, K. Christensen, and J. W. Vaupel. Rise, stagnation, and rise of Danish women's life expectancy. *Proceedings of the National Academy of Sciences*, 113 (15):4015–4020, 2016.
- W. Lutz and S. Kc. Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554):2779–2791, 2010.
- M. F. MacDorman, D. L. Hoyert, and T. Mathews. Recent Declines in Infant Mortality in the United States, 2005-2011. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2013.
- J. P. Mackenbach. Nordic paradox, Southern miracle, Eastern disaster: persistence of inequalities in mortality in Europe. The European Journal of Public Health, 27(suppl_4):14–17, 2017.
- M. Marmot. Social determinants of health inequalities. The Lancet, 365(9464):1099-1104, 2005.
- S. Mazzuco, B. Scarpa, and L. Zanotto. A mortality model based on a mixture distribution function. *Population Studies*, 72(2):191–200, 2018.
- M. Meilă. Comparing clusterings an information based distance. Journal of Multivariate Analysis, 98:873–895, 2007.
- P. Müller, F. Quintana, and G. L. Rosner. A product partition model with regression on covariates. Journal of Computational and Graphical Statistics, 20(1):260–278, 2011.
- OECD. Data warehouse. 2024. doi: https://doi.org/https://doi.org/10.1787/data-00900-en.
- J. Oeppen and J. W. Vaupel. Broken limits to life expectancy. Science, 296(5570):1029-1031, 2002.
- G. L. Page, F. A. Quintana, and D. B. Dahl. Dependent modeling of temporal sequences of random partitions. Journal of Computational and Graphical Statistics, 31(2):614–627, 2022.
- F. Pavone, S. Legramanti, and D. Durante. Learning and forecasting of age-specific period mortality via b-spline processes with locally-adaptive dynamic coefficients. *The Annals of Applied Statistics*, 18(3):1965–1987, 2024.
- A. E. Raftery, J. L. Chunn, P. Gerland, and H. Ševčíková. Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801, 2013.
- S. Schnürch, T. Kleinow, and R. Korn. Clustering-based extensions of the common age effect multi-population mortality model. *Risks*, 9(3):45, 2021.
- S. Scognamiglio. A multi-population locally-coherent mortality model. In Mathematical and Statistical Methods for Actuarial Sciences and Finance, pages 423–428. Springer, 2022.
- C. Tang, H. L. Shang, and Y. Yang. Clustering and forecasting multiple functional time series. The Annals of Applied Statistics, 16(4):2523-2553, 2022.
- J. Vallin and F. Meslé. Convergences and divergences in mortality: a new approach of health transition. Demographic Research, 2:11–44, 2004.
- J. W. Vaupel, Z. Zhang, and A. A. van Raalte. Life expectancy and disparity: an international comparison of life table data. *BMJ Open*, 1(1):e000128, 2011.
- S. Wade and Z. Ghahramani. Bayesian cluster analysis: point estimation and credible balls. Bayesian Analysis, 13(2):559–626, 2018.
- C. K. Williams and C. E. Rasmussen. *Gaussian Processes for Machine Learning*. MIT Press Cambridge, MA, 2006.
- C. Wilson. Understanding global demographic convergence since 1950. Population and Development Review, 37 (2):375–388, 2011.
- J. S. Wong, J. J. Forster, and P. W. Smith. Bayesian mortality forecasting with overdispersion. Insurance: Mathematics and Economics, 83:206-221, 2018.
- S. H. Woolf and H. Schoomaker. Life expectancy and mortality rates in the United States, 1959-2017. JAMA, 322 (20):1996–2016, 2019.