Estimating demographic and global health indicators for multiple countries and periods in the context of missing data and data quality issues

introducing a class of temporal models for multiple populations to facilitate model comparison

Herb Susmann[a], Monica Alexander[b], Leontine Alkema[a]

[a]University of Massachusetts Amherst, [b]University of Toronto

April 29, 2021

Resources

- Annotated slides: http://herbsusmann.com/paa2021
- Preprint: https://arxiv.org/abs/2102.10020

Statistics > Methodology

[Submitted on 19 Feb 2021]

Temporal models for demographic and global health outcomes in multiple populations: Introducing a new framework to review and standardize documentation of model assumptions and facilitate model comparison

Herbert Susmann, Monica Alexander, Leontine Alkema

There is growing interest in producing estimates of demographic and global health indicators in populations with limited data. Statistical models are needed to combine data from multiple data sources into estimates and projections with uncertainty. Diverse modeling approaches have been applied to this problem, making comparisons between models difficult. We propose a model class, Temporal Models for Multiple Populations (TMMPs), to facilitate documentation of model assumptions in a standardized way and comparison across models. The class distinguishes between latent trends and the observed data, which may be noisy or exhibit systematic biases. We provide general formulations of the process model, which describes the latent trend of the indicator of interest. We show how existing models for a variety of indicators can be written as TMMPs and how the TMMP-based description can be used to compare and contrast model assumptions. We end with a discussion of outstanding questions and future directions.

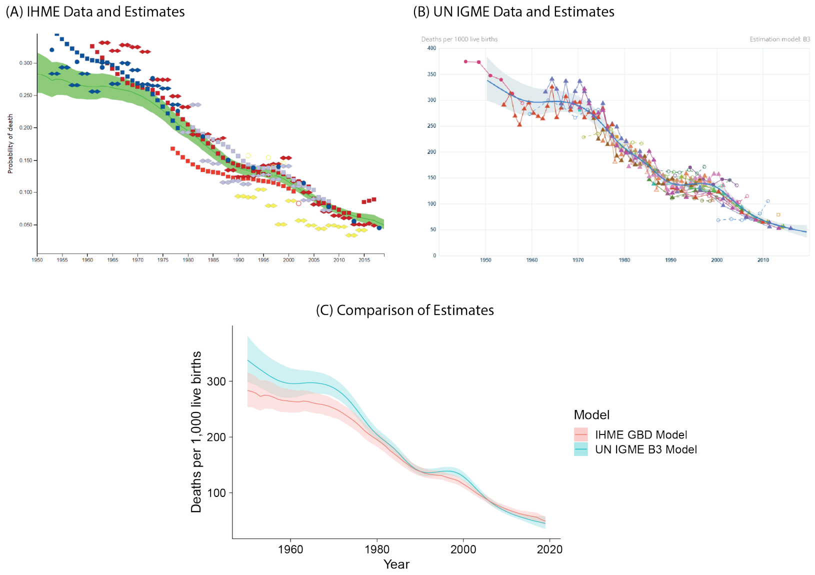
Background

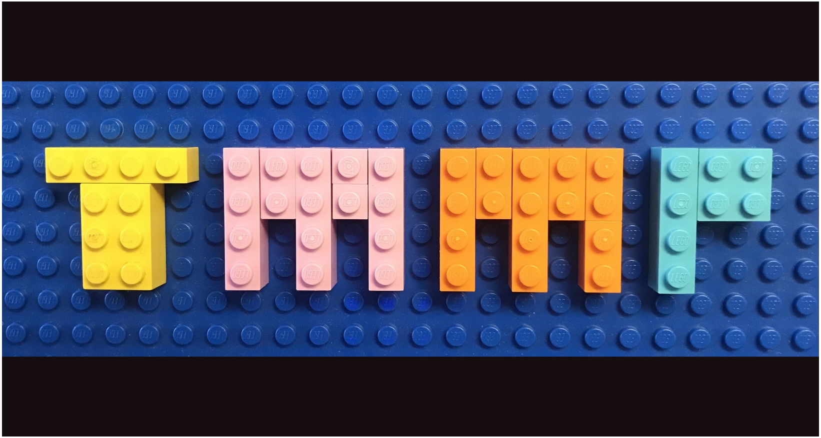
- There is interest in modeling demographic and health indicators in order to measure progress towards international goals.
- Data availability and quality are varied.
- Many statistical models have been created to provide estimates and projections.
- Comparing across models can be difficult.
- Proposed overarching model class: Temporal Models for Multiple Populations (TMMPs)

Case Study: U5MR

UMass Amherst

Under-five Mortality Rate Estimates in Senegal, 1950-2019

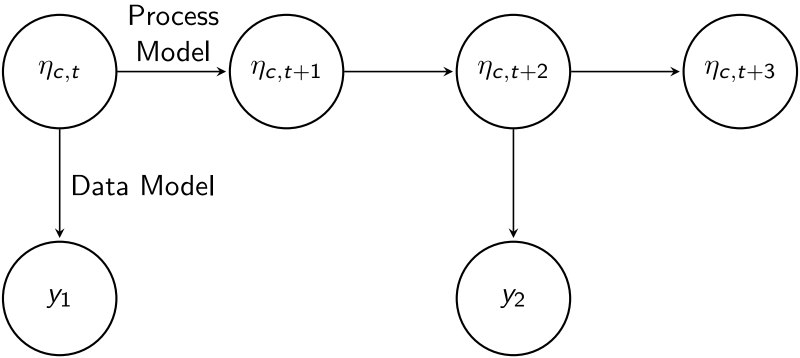




Modeling Framework

- True value of indicator: $\eta_{c,t}$
- *Process model* describes evolution of $\eta_{c,t}$
  - Covariates
  - Systematic trends
- Observed data y_i, with associated properties $c[i]$, $t[i]$, $s[i]$, ...
- *Data model* describes relationship between y_i and $\eta_{c[i],y[i]}$

Modeling Framework



Process Model

UMassAmherst

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

Covariate component

UMassAmherst

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- Regression function for incorporating covariates
- Example: Neonatal mortality rate model [Alexander and Alkema, 2018]

$$g_2(X_{c,t}, \beta_c) = \beta_{0,c} + \beta_1 \cdot (\log(X_{c,t}) - \log(\beta_2))1(X_{c,t} > \beta_2)$$

Systematic component

UMassAmherst

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- Parametric function for modeling systematic temporal trends
- Example: logistic curves for modeling adoption of modern family planning methods [Cahill et al., 2018]

Offset

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- The offset term incorporates external information, for example from a separate modeling step
- Example: IHME U5MR model uses an offset derived from the smoothed residuals of a separate mixed-effects model

Smoothing component

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- The smoothing component allows data-driven deviations from the other components, while still enforcing smoothness
- Many choices B-splines, Gaussian processes, AR(p), RW(p), spatial smoothing (ICAR), ...
- Model class introduces some additional structure to help understand smoothing

Smoothing component

UMassAmherst

- Define $\boldsymbol{\epsilon}_c = [\epsilon_1, \cdots, \epsilon_T]$
- Smoothing model defined as

$$\boldsymbol{\epsilon}_c = \boldsymbol{B}_c \boldsymbol{\alpha}_c$$

where $\boldsymbol{B}_c$ is a full rank matrix, and

$$\triangle_r \boldsymbol{\alpha}_c \sim N(\mathbf{0}, \boldsymbol{\Sigma}_c)$$

with $\boldsymbol{\Sigma}_c$ defined via an autocovariance function s.

Smoothing component

UMassAmherst

- AR(1), $r = 0$,

  $$s(t_1, t_2) = \kappa^2 \rho^{|t_1 - t_2|}$$

- RW(2), $r = 2$,

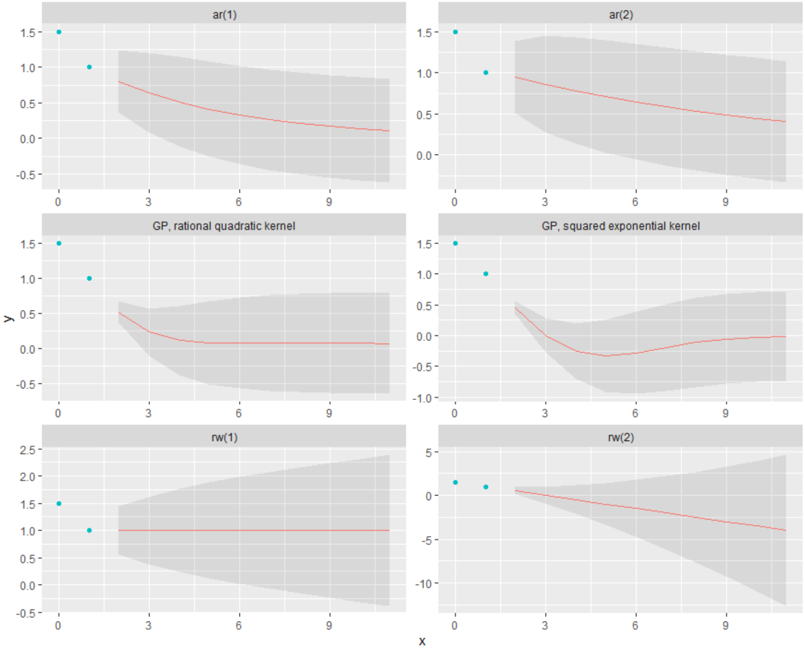
  $$s(t_1, t_2) = \kappa^2 I(t_1 = t_2)$$

- Matérn, $r = 0$,

  $$s_{\text{Matérn}}(t_1, t_2) = \kappa^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|t_1 - t_2|}{\ell} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{|t_1 - t_2|}{\ell} \right)$$

Smoothing component

UMassAmherst



Estimation

- Each component introduces many unit specific parameters that need to be estimated.
- Hierarchical modeling is a way to share information between units.
  - Example: hierarchically model county-specific regression coefficients $\beta_{k,a}$ [Alexander et al., 2017]
- Models can be fit with full Bayesian inference (JAGS, Stan), approximate Bayesian inference (INLA), or with frequentist methods (Template Model Builder)

UN-IGME U5MR Model

- Process model [Alkema and New, 2014]

$$\log\left(\eta_{c,t}\right) = \underbrace{g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{\epsilon_{c,t}}_{\text{smoother}}$$

- Systematic component:

$$g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c) = \alpha_{c,0} + \alpha_{c,1}(t - t_c^*)$$

- Smoothing component: cubic B-splines with K_c knots per country and RW(2) process on spline coefficients

$$\boldsymbol{\epsilon}_c = \boldsymbol{B}_c \boldsymbol{\delta}_c,$$

where after two levels of differencing $\boldsymbol{\delta}_c$ is normally distributed with mean zero:

$$\Delta_2 \boldsymbol{\delta}_c \sim N(\mathbf{0}, \sigma^2_{\delta,c} \boldsymbol{I}).$$

IHME U5MR Model

- Process model [Dicker et al., 2018]

$$\log_{10}(\eta_{c,t}) = \underbrace{g_2(\boldsymbol{X}_{c,t}, \boldsymbol{\beta}_c)}_{\text{covariate}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoother}}$$

- Covariate component:

$$g_2(\boldsymbol{X}_{c,t}, \boldsymbol{\beta}_c) = \exp\left[\beta_{c,1} \cdot \log(X_{c,t}^{LDI}) + \beta_{2,c} \cdot X_{c,t}^{EDU} + \beta_{3,c}\right] + \beta_{4,c} X_{c,t}^{HIV}$$

- Offset: adjusts covariate component using smoothed residuals from separate mixed-effects model
- Smoother: Gaussian process with no transformation $(\boldsymbol{B} = \boldsymbol{I}, \boldsymbol{\epsilon}_c = \boldsymbol{\delta}_c)$ and Matérn covariance function

$$\boldsymbol{\delta}_c \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_c).$$

Discussion

- Introduced Temporal Models for Multiple Populations (TMMPs)
- Model class makes a clear distinction between the *process model* and the *data model*
- Process model is split into building blocks: covariates, systematic trends, offsets, and smoothing models
- Facilitates comparisons between existing models
- Can be used as a framework for development of new models

References

Alexander, M. and Alkema, L. (2018).
Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model.
Demographic Research, 38:335–372.

Alexander, M., Zagheni, E., and Barbieri, M. (2017).
A Flexible Bayesian Model for Estimating Subnational Mortality.
Demography, 54(6):2025–2041.

Alkema, L. and New, J. R. (2014).
Global estimation of child mortality using a Bayesian B-spline Bias-reduction model.
The Annals of Applied Statistics, 8(4):2122–2149.

Cahill, N., Sonneveldt, E., Stover, J., Weinberger, M., Williamson, J., Wei, C., Brown, W., and Alkema, L. (2018).
Modern contraceptive use, unmet need, and demand satisfied among women of reproductive age who are married or in a union in the focus countries of the Family Planning 2020 initiative: a systematic analysis using the Family Planning Estimation Tool.
The Lancet, 391(10123):870–882.

Dicker, D., Nguyen, G., Abate, D., and other GBD 2017 Mortality Collaborators (2018).
Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017.
The Lancet, 392(10159):1684–1735.