A Primer on Small-Area Estimation

A Case Study on Mortality in Massachusetts Census Tracts

Eugenio Paglino

October 26, 2023

Department of Sociology, Population Studies Center, University of Pennsylvania

Small Area Estimation generally refers to **the estimation of a quantity of interest for a small subpopulation**. In demography, Small Area Estimation problems arise in many situations because **most demographic events** (birth, death, and migration) **are relatively rare**, especially when their occurrence is measured in specific age-ranges and periods.

For example, in 2021, the US mortality rates for age groups 1-4, 5-9, 10-14, and 15-19 are all below 1 every 1,000 residents (HMD, 2023). Which, given the US age-distribution, means we would need at least 15,000 residents to observe at least one death in each of these age-groups each year.

Given that about **1 every 3 counties in the US have fewer than 15,000 residents**, you see why techniques for modelling demographic quantities in these areas are crucial.

The issue with **zero death counts** in small populations is that they **provide little information about the underlying mortality rate**. They give us a sense that mortality is low but make it very difficult for us to say where exactly the rate should be.

Clearly, a direct estimation approach would fail in this case because a mortality rate equal to zero is effectively impossible. Even if tiny, the probability of dying over a positive time interval should always be positive.

In other words, **with a large enough population we should always observe a strictly positive number of deaths**.

Our beloved demographic methods textbook (Preston et al., 2001) discusses few statistical models for mortality. The Gompertz model (Gompertz, 1825), the Brass relational model (Brass, 1971), and the Lee-Carter model for forecasting mortality (Lee and Carter, 1992).

With the exception of the Lee-Carter model, the other models are only introduced in the context of unreliable data (together with model life-tables). Indeed, the book gives the impression that direct estimation is always preferred when the data is deemed reliable.

Reliable versus Unreliable Data

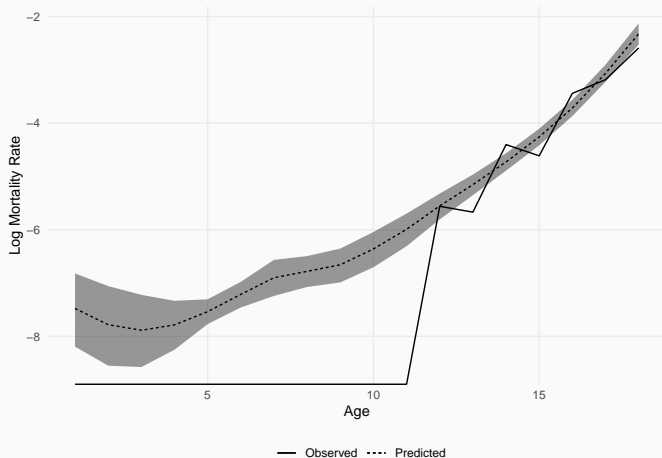However, models are useful even when the data is complete and reliable.

Figure 1: Direct vs. Model-Based Estimation

Hierarchical models (also called multilevel models, mixed effects models, or random effects models - don't ask me why all these names) can often provide a solution to estimation issues arising with sparse data.

From the frequentist viewpoint, hierarchical models can be thought of as models in which the parameters themselves are random variables and have a distribution whose parameters we are trying to estimate.

If this sounds confusing it is likely because in the frequentist framework we make distributional statements about y but never about β (e.g. the rather absurd language about confidence intervals). If this does not sound right to you it's probably because you are thinking of $\hat{\beta}$, which is indeed a random variable but should not be confused with β.[1]

This is probably confusing but it may become clearer with an example (after one more slide of theory).

---

[1]This paragraph alone is a reason enough to switch over to a Bayesian approach.

Every Bayesian model has three components[2]:

- **Likelihood** $p(data|unknowns)$. A function showing the probability of observing the data, for any particular value of the unknowns. A way of summarizing the information about unknowns contained in the data.
- **Prior distribution** $p(unknowns)$. A probability distribution summarizing information about the unknowns, beyond what is contained in the data.
- **Posterior distribution** $p(unknowns|data)$. A probability distribution summarizing information about the unknowns, after combining information from the likelihood and prior.

Let's go to an example.

_____

[2]The excellent language describing each component is borrowed from Bryant and Zhang (2018)

A Demographer's Best Friend: The Poisson Model

A one-parameter Poisson model:

$$Likelihood : p(y|\theta) = \prod_i \frac{\theta^{y_i} e^{-\theta}}{y_i!} = \frac{\theta^{\sum_i y_i} e^{-n\theta}}{\prod_i y_i!} \tag{1}$$

$$Prior : p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \tag{2}$$

$$Posterior : p(\theta|y) \propto p(y|\theta)p(\theta|\alpha, \beta) \tag{3}$$

$$\propto \frac{\theta^{\sum_i y_i} e^{-n\theta}}{\prod_i y_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \tag{4}$$

$$\propto \theta^{\sum_i y_i} e^{-n\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \tag{5}$$

$$\propto \theta^{\sum_i y_i + \alpha - 1} e^{-(n\theta + \beta)} \tag{6}$$

$$\implies \theta|y \sim Gamma\left(\sum_i y_i + \alpha, n + \beta\right) \tag{7}$$

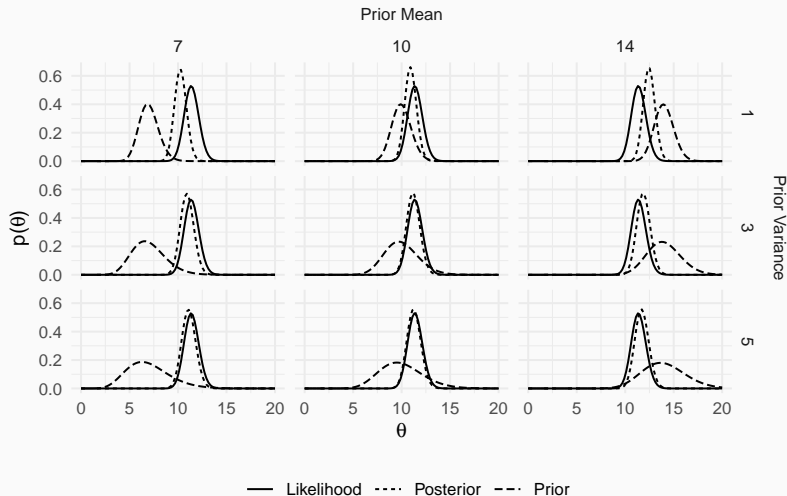You can run this example with the file *BDAOverview.Rmd*.

Figure 2: Bayesian Inference for the One-Parameter Poisson Model

Now we are going to work through a more complex example. Open the following files in RStudio:

- *hierarchicalRegression.Rmd*
- *regressionModel.stan*
- *regressionModelFE.stan*
- *regressionModelH.stan*

Now that we have gotten a sense of what hierarchical models do, we can get to a real application to estimate tract-level life-expectancies in Massachusetts.

- Deaths: Tract-Level death counts by five-year age groups for the period 2017-2021.
- Population: Population counts by five-year age groups from the 2020 Census.

I confidentialized the mortality data by applying random rounding to base three. The exposure data (population) is not ideal but no yearly population estimates are produced by the Census Bureau at the tract level.
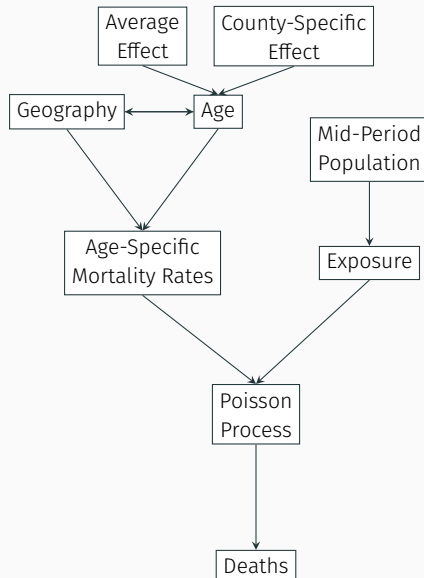
Figure 3: Analytical Framework

Denoting deaths for age group a, and tract s with $D_{a,s}$, the corresponding population with $P_{a,s}$, and the corresponding mortality rate with $m_{a,s}$, we have:

$$D_{a,s} \sim Poisson(m_{a,s} \cdot P_{a,s}) \tag{8}$$

$$log(m_{a,s}) = \beta_0 + \beta_1^A \cdot AC_{1,a} + (\beta_2^A + \beta_{2,s}^A) \cdot AC_{2,a} + (\beta_3^A + \beta_{3,s}^A) \cdot AC_{3,a} + \beta_s \tag{9}$$

$$\beta_0, \beta_1^A, \beta_2^A, \beta_3^A \sim Unif(-\infty, +\infty) \tag{10}$$

$$\beta_{2,s}^A \sim Normal(0, \sigma_2^A) \tag{11}$$

$$\beta_{3,s}^A \sim Normal(0, \sigma_3^A) \tag{12}$$

$$\beta_s \sim Normal(0, \sigma^S) \tag{13}$$

This model is indebted to Alexander et al. (2017). More complex models are discussed in Bryant and Zhang (2018), which also provides a broad overview of applications of Bayesian statistics to demography.

The Model: Age

Age is modelled through tract-specific coefficients on three vectors (AC_1, AC_2 and AC_3) representing common underlying features of the age-mortality curves. I obtained the three vectors as the first three principal components of county-level age-mortality curves.
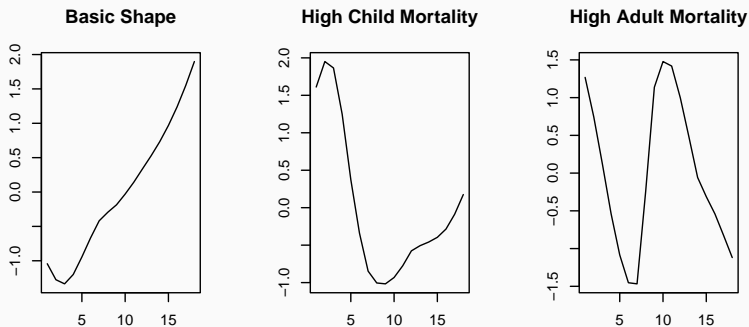

Figure 4: First Three Principal Components of County-Level Age-Mortality Curves

Geography is modelled with a set of county-specific intercepts assumed to come from the same Normal distribution. This hierarchical structure induces global smoothing, especially for less populated counties.

The model combines all the components we have seen and is estimated (with several implementation challenges) in Stan (Stan Development Team, 2023).

$$D_{a,s} \sim Poisson(m_{a,s} \cdot P_{a,s})$$
$$log(m_{a,s}) = \beta_0 + \beta_1^A \cdot AC_{1,a} + (\beta_2^A + \beta_{2,s}^A) \cdot AC_{2,a} + (\beta_3^A + \beta_{3,s}^A) \cdot AC_{3,a} + \beta_s$$

The result are 4,000 samples from the posterior distribution of each parameter that are used to construct posterior samples for the log mortality rates of all 5,187 data points.
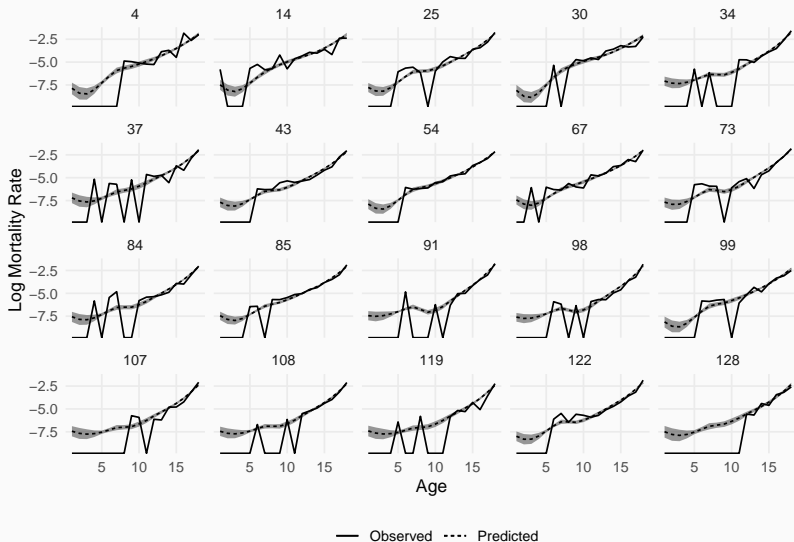
Figure 5: Comparing Estimated and Observed Mortality Rates by Age
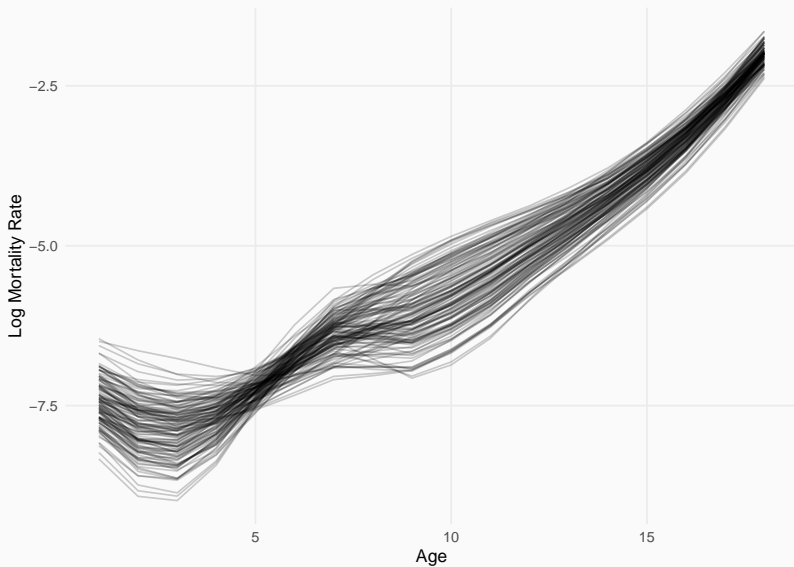
Figure 6: Age-Mortality Curves
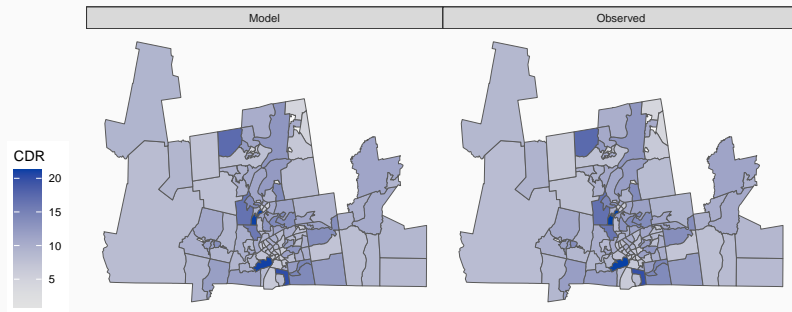
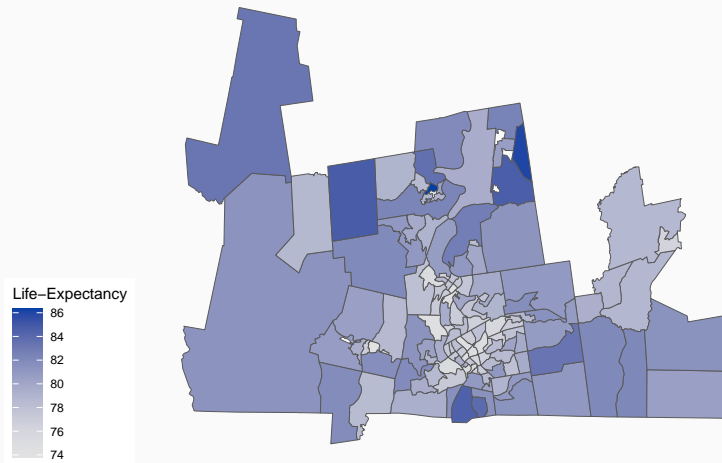**Figure 7:** Crude Death Rates by Tract

Figure 8: Life Expectancies by Tract

The good stuff:

- Models for mortality rates are a very useful tool in a demographer's toolkit.
- They are useful even when the data is complete and reliable and make it much easier to deal with uncertainty in a systematic way (sorry Chiang (1984), it was good until it lasted).
- (personal opinion) The Bayesian approach (especially with pure Stan) forces you to think hard about the model and demands a very good understanding of what you are doing (which is not always the case with frequentist procedures).

The caveats:

- Care should be used when selecting an appropriate model. Are we capturing all important features of the data? (Not always easy to answer).

- The Achilles's heel of Bayesian inference is that it is computationally intensive and hard to scale when the number of data points increases (Stan is the fastest of the MCMC approaches but approximation approaches like INLA or Template Model Builder are generally faster). See Dwyer-Lindgren et al. (2016) for an example of county-level modeling using Template Model Builder.

- Bayesian approaches are not the only option. The National Center for Health Statistics undertook a massive (and surprisingly ignored) effort to produce tract-level life-expectancy estimates using a frequentist (but poorly documented) approach (Arias et al., 2018).

References

Alexander, M., Zagheni, E., and Barbieri, M. (2017). A flexible bayesian model for estimating subnational mortality. *Demography*, 54(6):2025–2041.

Arias, E., Escobedo, L. A., Kennedy, J., Fu, C., and Cisewski, J. A. (2018). Us small-area life expectancy estimates project: Methodology and results summary.

Brass, W. (1971). On the scale of mortality. *Biological aspects of demography*.

Bryant, J. and Zhang, J. L. (2018). *Bayesian demographic estimation and forecasting*. CRC Press.

Chiang, C. L. (1984). Life table and its applications. In *Life table and its applications*, pages 316–316.

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Kutz, M. J., Huynh, C., Barber, R. M., Shackelford, K. A., Mackenbach, J. P., van Lenthe, F. J., et al. (2016). Us county-level trends in mortality rates for major causes of death, 1980-2014. *Jama*, 316(22):2385–2401.

Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.

HMD (2023). Human mortality database. max planck institute for demographic research (germany), university of california, berkeley (USA), and french institute for demographic studies (france). available at www.mortality.org.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Preston, S., Heuveline, P., Guillot, M., et al. (2001). Measuring and modeling population processes. *Wiley-Blackwell*.

Stan Development Team (2023). Stan modeling language users guide and reference manual, 2.32.