Gaussian Process Models for Mortality Improvement Factors

Longevity 12

Jimmy Risk

Dept of Statistics & Applied Probability UC Santa Barbara

September 29 2016

Joint with M. Ludkovski (UCSB) and H. Zail (Elucidor)

Mortality modeling

	Male			Female		
Year	Age	Exposure	Mortality	Age	Exposure	Mortality
2011	50	2,211,607	0.005003	50	2,279,824	0.003161
2011	64	1,661,474	0.014508	64	1,826,662	0.008898
2011	74	857,432	0.032229	74	1,032,934	0.021906
2011	84	411,265	0.085986	84	624,921	0.064030

Excerpt of CDC mortality data to compare exposures and mortality rates over ages and gender. *Exposure* is the estimated mid-year number of individuals at given age. *Mortality* is the observed proportion of the population deceased during the year.

Objectives:

- Smooth observed mortality experience
- Forecast unobserved mortality (future years, additional ages)
- Quantify uncertainty (data-driven vs model-driven)
- Interpret mortality evolution, especially mortality improvement factors

Main Message

- Present a new *statistical* framework based on Gaussian Process regression
- Coherently deal with smoothing and extrapolation
- Nonparametric method
- Bayesian paradigm to quantify model/predictive uncertainty
- Built in tools to analyze mortality improvement
- Believe this is a viable alternative to traditional APC/Lee-Carter frameworks
- Amenable to many extensions that will be treated in future works

US Mortality Dataset

- 2-D table indexed by Age and Year: $x = (x_{ag}, x_{yr})$
- Raw deceased count D_{ij}; Exposures E_{ij}
- Latent log-mortality surface $f(x)$
- Noisily observed in raw rates $Y \equiv \mu_{ij} = \log \frac{D_{ij}}{E_{ij}}$
- $Y(x) = f(x) + \varepsilon$ where $Var(\varepsilon(x)) = \sigma^2(x)$
- CDC dataset: years 1999-2014, ages 50-84
- (Ignore old data or Young ages since focus on annuity/pension plan projections)

Statistical Framework

- Treat the true mortality surface $f \in \mathcal{H}$ as a random function
- Specify prior distribution and then use Bayesian updating
- Output is a posterior distribution: provides both the point estimate and credible bands $p(f|\mathcal{D}) \propto p(\boldsymbol{y}|f, \boldsymbol{x})p(f) = \{likelihood\} \cdot \{prior\}$
- Covariance structure: knowing mortality at x will greatly influence mortality at “neighboring” x’s: the mortality rate for a 60 year old in 2015 will be closer to that of a 61 year old in 2016, than that of a 20 year old in 2050

Gaussian Processes

- f is a realization of a Gaussian random field with a covariance structure defined by C, function space $\mathcal{H}_C = span(C(\cdot, x) : x \in \mathcal{X})$
- $C(x, x') := \mathbb{E}[f(x)f(x')]$ controls the spatial smoothness
- Squared-Exponential kernel

$$C(x, x') = \eta^2 \exp\left(-\frac{(x_{ag} - x'_{ag})^2}{2\theta_{ag}^2} - \frac{(x_{yr} - x'_{yr})^2}{2\theta_{yr}^2} \right)$$

- Lengthscales θ's and fluctuation scale η.
- Prior $p(\boldsymbol{f}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$
- Observation likelihood $p(\boldsymbol{y}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{f}, \Sigma)$
- The posterior conditional on $(x, y)^{1:N}$ is also Gaussian $f(x)|\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{N}(m_*(x), s_*^2(x))$

$$m_*(x) = \vec{c}(x)^T (\mathbf{C} + \Sigma)^{-1} \vec{y}$$

$$s_*(x, x') = K(x, x') - \vec{c}(x)^T (\mathbf{C} + \Sigma)^{-1} \vec{c}(x')$$

- $C_{ij} = C(x^i, x^j)$, $\Sigma = diag(\sigma^2(x^i))$, $c_i = C(x, x^i)$

GP details

- Mean function $m(x)$: captures the Age-shape of mortality and long-term Year trend
- Beyond the dataset $m_*(x) \to m(x)$ revert to the prior
- Lengthscales θ control the above transition (akin to bandwidth in kernel regression)
- Observation noise: more correct is to use binomial $D_{ij} \sim Bin(E_{ij}, \mu_{ij})$
- Empirically there is overdispersion (partly because $e^{f_{ij}}$ is unknown)
- For fitting purposes can replace with a constant “nugget” σ that is estimated by MLE

Fitting a GP

- Must specify the kernel family (other popular choices include Matern-5/2, et cetera)
- Need to know the kernel hyperparameters – η, θ's, et cetera.
- Use MLE (nonlinear optimization problem) or cross-validation
- We used `DiceKriging` package in `R` – off-the-shelf use

Bayesian GP

- To quantify model uncertainty specify priors for the kernel hyper-parameters
- Hierarchical approach for determining the covariance matrix: "automatic relevance determination"
- Posterior is a mixture of Gaussians
- Sample from the posterior using MCMC
- Care is required due to poor mixing, eg. Hamiltonian MCMC
- Implemented in `Stan`

Comparison to Lee-Carter

- LC postulates that $\mu_{ij} = \alpha_i + \beta_i \kappa_j + \varepsilon_{ij}$
- Two-step: first fit the factors α, β, κ using *the entire* dataset
- Then build a time-series model for time-dependent factors to make forecasts
- Postulates a priori the structural dependence
- Smoothing is via a point estimator, no credible bands

Estimated Mortality Covariance Structure

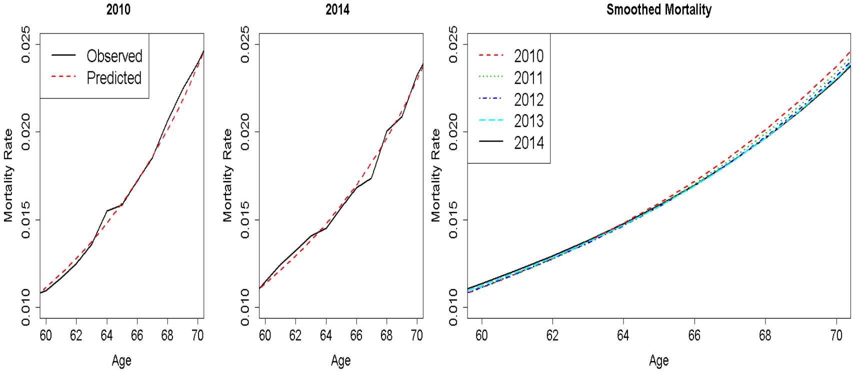
	DiceKriging	STAN		
	MLE	MAP	MCMC Mean	MCMC 80% Posterior CI
θ_{ag}	15.8384	14.9320	10.3580	(4.8976, 16.3939)
θ_{yr}	15.5308	14.4895	24.6674	(12.8976, 38.2304)
η^2	1.8468	1.2372	1.8862	(0.7618, 3.5324)
σ^2	2.808e-04	2.752e-04	2.745e-04	(2.5031e-04, 2.988e-04)
β_0	-3.8710	-3.8277	-3.7966	(-4.5986, -3.0185)

Comparing hyperparameter estimates for MLE (`DiceKriging`) and MAP (STAN) along with MCMC summary statistics. The GP is fitted to all data and uses squared-exponential covariance kernel. Priors for θ_{ag}, θ_{yr}, and η^2 Log Normal$(0, 1)$; prior for σ^2 is Half-Normal$(0, 0.2)$; prior for β_0 is Normal$(-4, 5)$ (after standardizing data).

Mean Function

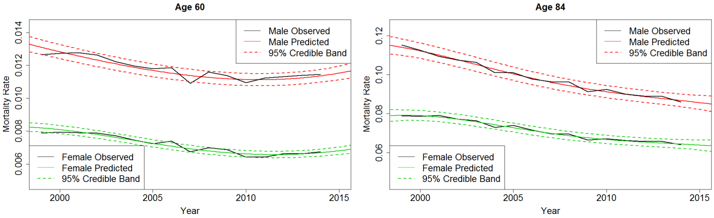
- The shape of f is a blend of the prior and the influence of the data
- At edges/beyond the dataset, f is driven by the prior
- Important to make the latter reasonable for extrapolation. In-sample effects are secondary
- In age, log-mortality is increasing (super-linearly):
  $m(x) = \beta_0 + \beta_1^{ag} x_{ag} + \beta_1^{yr} x_{yr}$ or
  $m(x) = \beta_0 + \beta_1^{ag} x_{ag} + \beta_1^{yr} x_{yr} + \beta_2^{ag} x_{ag}^2$
- In year, log-mortality is decreasing

Mortality Smoothing



Mortality rates for Males aged 60-70 during the years 2010-2014. Raw vs. estimated smoothed mortality curves. Models are fit to 1999–2014 CDC data for Ages 50–84 (All data).

Mortality over time

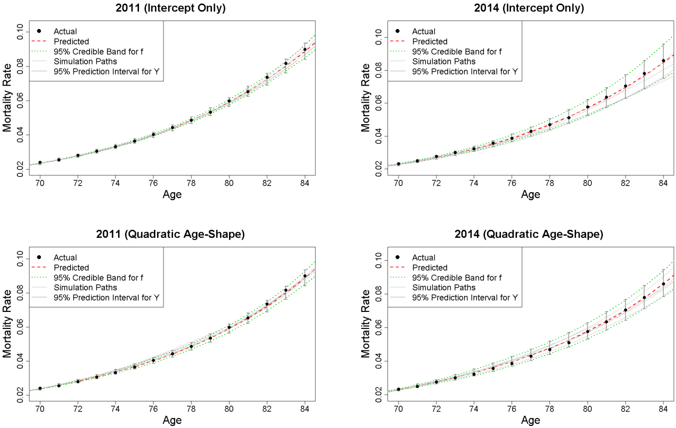


Mortality rates for Males (top) and Females (bottom) aged 60 and 84 over time. The plots show **raw mortality** rates for years 1999-2014, as well as predicted mean of the smoothed mortality surface and its 95% credible band. Models are fit to 1999–2014 CDC data for Ages 50–84 (All data).

In-Sample Smoothing:

- Results are stable across different implementations of GPs
- In-sample prediction is data-driven, priors/covariance structure/observation noise plays secondary role

Mortality Credible Bands



Mortality rate prediction for years 2011 and 2014 and ages 71-84. Model is fit with Subset II data with intercept-only (top) and quadratic-shape (bottom) mean functions and squared-exponential kernel. "Simulation paths" refer to a simulated path of the true f. Credibility bands are for the true mortality surface f; vertical intervals are for predicted observable mortality experience.

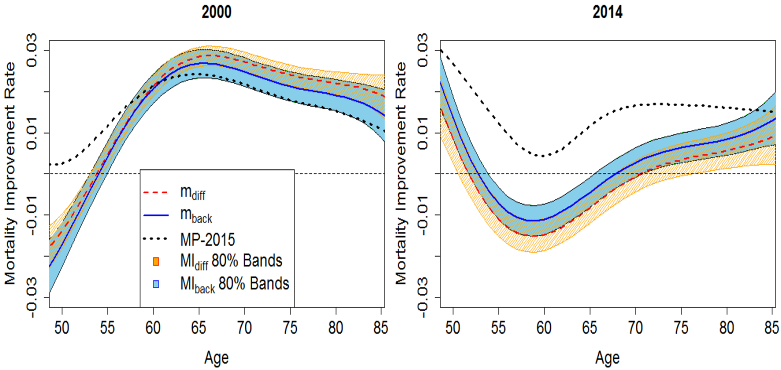
Discussion

- Can sample full trajectories for future mortality evolution (not just marginally, but the full fitted dependence in Ages or Years)
- Above is very useful for risk management (eg assessing VaR of annuity portfolio)
- Intrinsically determines the range of meaningful (data-driven) forecasts. We find that can forecast $\sim 5-7$ years into the future
- The GP model is updateable: directly shows the impact of new/additional data on the projections
- Choose to model log-mortality, but can also model other quantities

Improvement Factors

- MI: change in f as a function of x_{yr}
- Traditional is backward-looking: $f(ag, yr) - f(ag, yr - 1) =: MI_{back}$
- MI_{back} remains a GP
- Instantaneous: $MI_{diff} := \partial_{yr} f(ag, yr)$
- Can analytically differentiate the latent log-mortality surface
- $\partial_x f$ is again a GP!
- Can quantify uncertainty on improvement factors – not possible in other approaches

Improvement Factors



Estimated male mortality improvement using the differential GP model (instantaneous improvement) and the YoY improvement from the original GP model. We show the means and 80% uncertainty bands for MI^{GP}_{diff} and MI^{GP}_{back} for males aged 50–84 and years 2000 & 2014. Models used are fit to All data.

Declining Longevity?

- Results strongly suggest that longevity is now decreasing at Ages 55-70
- Diverges from the MP-2015 scales that continue to assume improvement
- Results are non-parametric and data-driven, so arguably as objective as it gets
- Significant implications for valuation of annuities and other contracts.
- Improvement is highly age-dependent

Data Relevance

- For actuaries, central issue is longevity risk: middle/older ages, today/future
- Hence, we build our models for Age 50+/1999+
- The mortality surface is non-stationary, so more distant data is detrimental (?)
- Our analysis suggests that looking at more than 20 years of data is irrelevant
- Perhaps need further segmenting over ages but seek balance over credibility/over-fitting/non-stationarity

Next Steps

- Monotonicity constraint
- Sub-populations
- Multiple populations
- Modeling by cause of death

Next Steps

- Monotonicity constraint
- Sub-populations
- Multiple populations
- Modeling by cause of death

THANK YOU!

References

Williams, C. K. and Rasmussen, C. E. 2006.
Gaussian processes for machine learning, the MIT Press.

Girosi, F. and King, G. 2008.
Demographic forecasting, Princeton University Press.

Currie, I. D. 2016.
On fitting generalized linear and non-linear models of mortality.
Scandinavian Actuarial Journal, 2016(4):356–383.

Dokumentov, A. and Hyndman, R. J. 2014.
Bivariate data with ridges: two-dimensional smoothing of mortality rates.
Technical report, Working paper series, Monash University.

Roustant, O., Ginsbourger, D., Deville, Y., et al. 2012.
Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization.
Journal of Statistical Software, 51(1):1–55.

M. Ludkovski, J. Risk, H. Zail
Gaussian Process Models for Mortality Rates and Improvement Factors
https://arxiv.org/abs/1608.08291
Submitted to North American Actuarial Journal