Gaussian Process Models for Incremental Loss Ratios

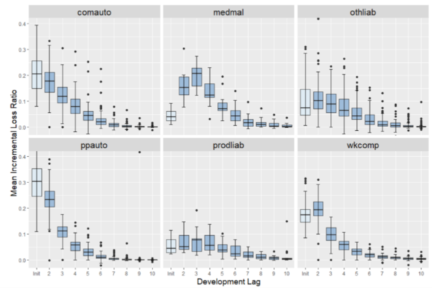
OICA Conference
April 29, 2020

Mike Ludkovski

Dept of Statistics & Applied Probability UC Santa Barbara
joint with H. Zail (Elucidor). Supported by CAS CKER grant

Loss Development Triangles

- Fundamental Challenge: completing the square
- Upper left triangle is the training data
- Lower right triangle: extrapolation
- Objectives:
  - Setting reserves: Expected value of Ultimate claims (Sum of column 10)
  - Allocating risk capital: Distribution of Ultimate claims
  - Expected value of successive step-ahead claims for cash flow projections & asset / liability management
- Six business lines from 200 triangles covering 1988–2006 in the NAIC database of Meyers & Shi (2011)



GP Model for Incremental Loss Ratios

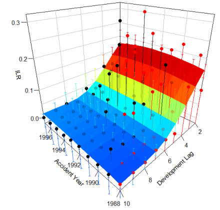
- Each "cell" in the triangle is treated as a data point: $x^i = (\mathrm{AY}^i, \mathrm{DL}^i)$ indexed by Accident Year p and Development Lag q
- Response is $L^i =$ incremental loss ratio $= I_{p,q}/P_p$
- $I_{p,q}$ is incremental loss, P_p is the pure premium paid, CC are cumulative losses observed

$$I_{p,q} := \begin{cases} CC_{p,1} & q = 1; \\ CC_{p,q} - CC_{p,q-1}, & q = 2, \ldots, Q = 10. \end{cases}$$

- Observed L^i is noisy version of true ILR surface $f(x^i)$
- $\boxed{L^i = f(x^i) + \epsilon_q}$, where $\epsilon_q \sim N(0, \sigma_q^2)$ (lag-dependent variance)
- Predictive task: predict $f(x)$ or $L(x)$ for an arbitrary x – in-sample or extrapolate
- Point forecast $m_*(x)$ and credible band $[\underline{m}_*(x), \bar{m}_*(x)]$
- Multivariate forecasts + future *scenarios* $L(\tilde{x}_{1:m})$

Training Dataset in red: $\mathcal{D} = (x^{1:n}, L^{1:n})$

Fitted $f(\cdot)$ is the response surface



Gaussian Processes

- Treat the true response surface f as a Gaussian random field w/prior $f \sim GP(\boldsymbol{m}(\mathbf{x}), \boldsymbol{C}(\mathbf{x}, \mathbf{x}))$
- Mean function $m(x^i) = \mathbb{E}[f(x^i)]$
- Covariance $C(x^i, x^j) = \mathbb{E}[(f(x^i) - m(x^i))\,(f(x^j) - m(x^j))]$
- multivariate squared-exponential kernel

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

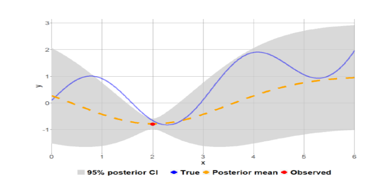
- Observation likelihood $p(\boldsymbol{L}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{L}|\boldsymbol{f}, \boldsymbol{\Sigma})$ (Gaussian conjugate!) w/$\boldsymbol{\Sigma} = diag(\sigma^2(x^i))$
- The posterior is Gaussian $f(x)|\mathcal{D} \sim \mathcal{N}(m_*(x), s_*^2(x))$ based on the multivariate Gaussian conditioning formula

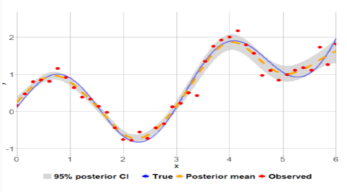
$$m_*(x) = \vec{c}(x)^T(\mathbf{C} + \Sigma)^{-1}\vec{L}$$
$$s_*(x, x') = C(x, x') - \vec{c}(x)^T(\mathbf{C} + \Sigma)^{-1}\vec{c}(x')$$

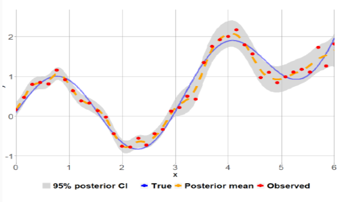
- $C_{ij} = \boldsymbol{C}(x^i, x^j)$, $\vec{c}_i = C(x, x^i)$

GP Modeling

- Capture the idea that f is learned from the data: specify prior distribution and then compute conditional distribution $p(f|\mathcal{D}) \propto p(\mathbf{L}|f, \boldsymbol{x})p(f) = \{likelihood\} \cdot \{prior\}$
- Covariance structure: response at x primarily influences responses at "neighboring" x's
- Given the kernel, the posterior is in closed-form
- Point estimate is $m_*(x)$
- Local prediction uncertainty $s_*(x)$
- Scenarios: samples from MVN posterior







Fitting a GP

- Specify the kernel family + learn the kernel hyperparameters $\Theta = (\eta, \theta$'s, etc.)
- Lengthscale θ controls correlation decay = spatial smoothness of $f(\cdot)$ and $m_*(\cdot)$
- Use MLE via the `DiceKriging` package in `R`;
- Bayesian hierarchical approach with priors on Θ via Hamiltonian MCMC implemented in `Stan`
- Mean Function
  - The shape of f is a blend of the prior mean/trend $m(x)$ and the influence of the data
  - At edges/beyond the dataset, f is driven by the prior
  - Use a mean function that is linear in AY and DL

Recap: GPs

- Spatial approach – no time-series
- Predictions resemble kernel regression (or smoothing splines)
- Non-parametric/data-driven
- Bayesian paradigm: Provides full probabilistic forecast + scenarios
- Can provide expert insights/structural assumptions via the prior mean/kernel
- Can modify observation likelihood like in GLMs
- Model Risk via Bayesian GP
- Good software that is getting even better – vast ML ecosystem

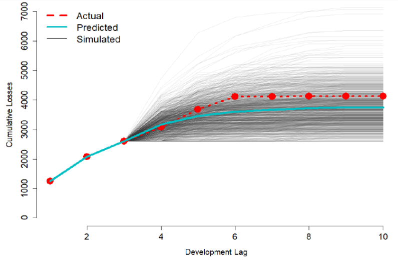
GP Model for Loss Development

- Work with incremental loss ratios which should be similar across triangles (see next talk by Q. Guo)
- Observed ILRs are noisy due to random fluctuations, actuarial concept of an underlying ILR surface is reasonable
- (Also tried analysis with Loss Development Factors $F_{p,q} := \frac{CC_{p,q}}{CC_{p,q-1}}$, leads to worse results)
- **Lally and Hartman** (IME 2018): GP model for cumulative losses

Uncertainty Quantification

Three layers of uncertainty:

- Extrinsic observation noise ϵ_q
- Intrinsic model uncertainty: credible band of the GP (given a covariance structure there are many random fields consistent with the data)
- Intrinsic correlation uncertainty: not sure about the triangle dependence pattern
- These multiple mitigate the underestimation of risk in $CC_{p,Q}$
- Mixture-of-Gaussians predictive distribution for $L_{p,q}$
- Generate consistent forecast for entire future $CC_{p,\cdot}$ (in Mack CL, future LDF's are iid)



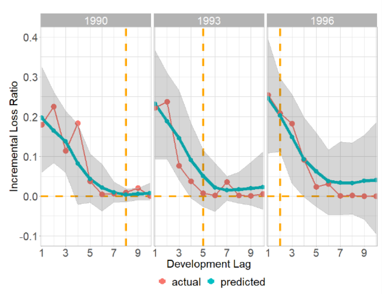
Structural Constraints

- As $q \to Q$, $f_{p,q} \to 0$ generally decreasing monotonically;
- Cumulative paid claims are increasing: $f_{p,q} \geq 0$;
- As $q \to Q$, $L_{p,q} \to 0$, generally monotonically, so there is little intrinsic and extrinsic uncertainty for long lags;
- As $q \to Q$, $Var(L_{p,q})$ decreases rapidly, so ILRs are highly volatile for short development lags, but have minimal variance for longer lags;

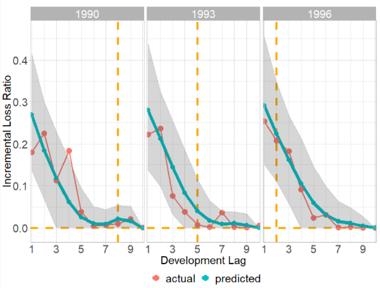
Implications for the GP model:

- $m_*(x) \to 0$ as $q \to Q$
- Intrinsic uncertainty (band tightens as q increases, $s_*(x) \to 0$
- Extrinsic uncertainty $\sigma^2(x)$ declines by DL and is right-skewed
- Log-transformation to enforce positivity does not work well; (LH proposed input warping)

Fitted GP Loss Ratios



Plain GP

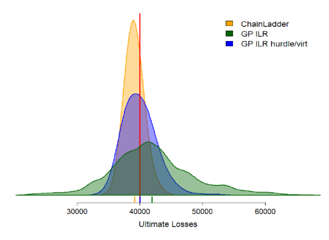


Enhanced GP

- Special observation model (hurdle: mixture of zero and positive): $L(x) \sim \mathcal{N}\left(m_*(x), \sigma^2(x)\right) \vee 0$
- Inferred hurdle probability h_q indicates how fast losses tend to develop
- Lag-Dependent observation noise σ_q
- Virtual observations at $q = 11$: augment training data with $L_{p,Q+1} = 0 \forall p$ with $\sigma_{Q+1} = 0$

Ultimate Loss Uncertainty Quantification

- Ultimate loss ratio $CC_{p,Q} = CC_{p,q} + P_p \left(\sum_{\ell=q+1}^{Q} L_{p,\ell}\right)$
- Best estimate reserves: RMSE of $CC_{p,Q}$
- Cash flow projections: RMSE of step-ahead cumulative loss ratio
- Risk Capital: Coverage Ratio (at 90% below)



Algorithm	Total RMSE	LR RMSE	Coverage	CRPS	NLPD	K-S
Mack CL	24726.2	0.049	0.509	388621	1305.6	0.306
Bootstrap CL	24895.7	0.052	0.544	389144	1232.5	0.292
ILR Plain	79812.0	0.138	0.965	1058856	1095.1	*0.146*
ILR Hurdle	41220.3	0.115	0.930	677361	1081.9	*0.106*
ILR Hurdle+Virt	42096.4	0.088	0.875	632531	1054.2	*0.071*

Percentile and Kolmogorov-Smirnov Tests

- Assessed the quality of the probabilistic forecast (predicted distribution of $CC_{p,Q}$ against observed ultimate losses across all triangles in a given business line)
- Measure CRPS (Continuous Ranked Probability Score) and NLPD (negative log probability density)
- Kolmogorov-Smirnov Test + distribution of percentile ranks

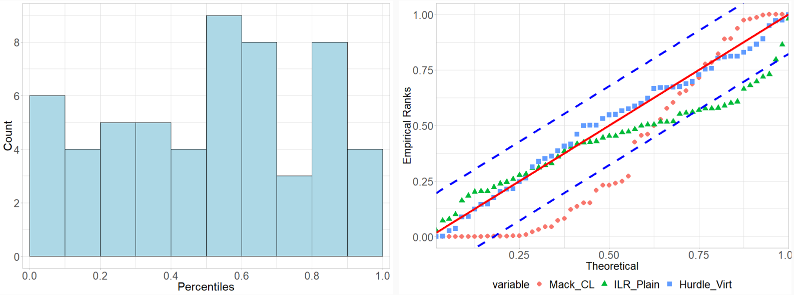


Figure 1: Left: Percentile rank of realized ultimate losses in terms of the predictive distribution of the ILR Hurdle+Virt model across 57 `wkcomp` triangles. Right: Kolmogorov-Smirnov test across three models for `wkcomp`. The dashed lines indicate the K-S test thresholds that are z_{KS} from the 45-degree line; only the Hurdle+Virt model passes the K-S test (stays between the dashed lines).

Take-Aways and Next Steps

- A rigorous spatial Bayesian perspective that is data-driven and fully stochastic
- Enables precise, nonparametric quantification of both extrinsic and intrinsic sources of uncertainty
- Significant improvement in coverage + rank percentile tests over competing approaches
- Document different dependence structures (ρ_{AY}, ρ_{DL}) across business lines
- Can extend to joint analysis over multiple triangles

$$C^{(SqExpMlt)}(x^i, x^j) := \eta^2 \exp(-\rho_{AY}^{-2}(\mathtt{AY}^i - \mathtt{AY}^j)^2 - \rho_{DL}^{-2}(\mathtt{DL}^i - \mathtt{DL}^j)^2) \cdot e^{-\rho_{Co}\cdot(1-\delta_{ij})}$$

- Requires scalable Stan/GP methods (Kronecker structure)

THANK YOU!

UCSB

References

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

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. Betancourt.
Robust Gaussian processes in Stan, 2017.
Online post at `https://betanalpha.github.io/assets/case_studies/gp_part1/part1.html`.

N. Lally and B. Hartman.
Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping.
Insurance: Mathematics and Economics, 82:124–140, 2018.

M. Ludkovski, H. Zail
Gaussian Process Models for Incremental Loss Ratios
Submitted, Revised Feb 2020, available by request.

Additional Plots

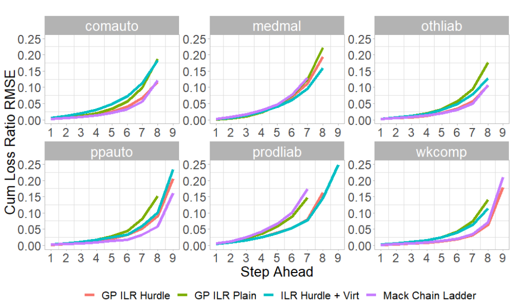


Figure 2: RMSE of cumulative loss ratios as a function of step-ahead *n* across the six business lines. Lags where RMSE exceeds 0.25 are clipped from the panels.

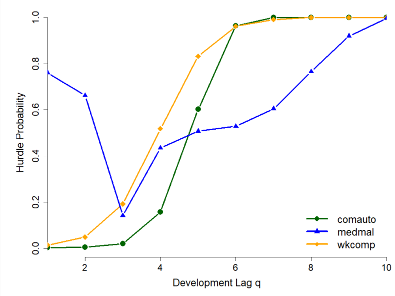


Figure 3: hurdle probability h_q. MCMC means across companies within the business line for three representative business lines.