Estimating Loss Reserves Using Hierarchical Bayesian Gaussian Process Regression with Input Warping

Brian Hartman (Brigham Young University)

Joint work with
Nathan Lally (Pratt and Whitney)

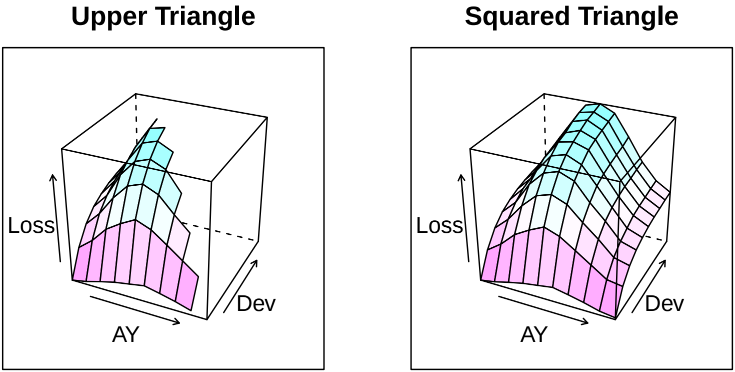
Advances in Predictive Analytics
01 December 2017
Paper and code available at `hartman.byu.edu`

State Farm's Workers' Compensation Incremental Claims Upper Triangle (Meyers and Shi, 2011)

	1	2	3	4	5	6	7	8	9	10
1988	22190	38644	24270	15047	8661	6155	3823	2768	1934	1557
1989	26542	51256	28609	16015	10937	5240	4430	2683	1646	
1990	32977	67517	34392	22872	11233	9074	4722	4973		
1991	38604	75824	42675	24219	16089	11393	4592			
1992	42466	83354	38956	24269	15332	9527				
1993	46447	70317	38133	24522	14257					
1994	41368	58976	31677	19060						
1995	35719	47497	28052							
1996	28746	37287								
1997	25265									

Development years across the columns, accident years down the rows.

State Farm Workers' Compensation Incremental Claims Triangle



Model

Model Goals

We would like to develop a model with

- Flexibility
- Parsimony
- Efficient computation
- Automation

A Gaussian process regression with input warping satisfies all these characteristics.

Gaussian Processes

- *Gaussian Process*: For all finite subsets of time points $t = \{t_1, t_2, ..., t_m\}$ in Ω, a process $\{Y_t\}_{t \in \mathcal{S}}$ is Gaussian if the joint distribution $(Y_{t_1}, Y_{t_2}, ..., Y_{t_m})$ is multivariate Gaussian.
- GPs are defined entirely by a mean function $M(\cdot) : \mathbb{R}^{n,p} \to \mathbb{R}^n$ and a covariance function $K(\cdot,\cdot) : \mathbb{R}^{n,p} \to \mathbb{R}^{n,n}$.
- A GP prior can be expressed by the following,

$$f \sim \mathcal{N}_n\left(M(X),\ K(X,X)\right)$$

for individual input pairs x and x'

$$\begin{aligned} m(x) &= \mathbb{E}\left[f(x)\right] \\ k(x,x') &= \mathbb{E}\left[(f(x) - m(x))\left(f(x') - m(x')\right)\right] \\ f(x) &\sim \mathcal{GP}\left(m(x),\ k(x,x')\right) \end{aligned}$$

Input Warping

One way to introduce non-stationarity into a stationary GP is through input warping. Each explanatory variable is normalized to [0,1] and then warped according to the following beta distribution

$$\omega_j(x_{ij}) = \int_0^{x_{ij}} \frac{u^{\alpha_j - 1}(1-u)^{\beta_j - 1}}{\mathrm{B}(\alpha_j, \beta_j)} du$$

Rather than define the warping explicitly, we place weakly informative priors on $\alpha_1, \beta_1, \alpha_2, \beta_2$.

$$\begin{aligned} \alpha_j &\sim \log \mathcal{N}(0,\ 0.5) \\ \beta_j &\sim \log \mathcal{N}(0,\ 0.5),\ \text{for } j \in \{1, 2\} \end{aligned}$$

Covariance Functions

$\mathcal{GP}_{\text{sq-exp iso}}$

$$\eta^2 \cdot \exp\left[-\psi(|w_1 - w_1'| + |w_2 - w_2'|)\right]$$

$\mathcal{GP}_{\text{sq-exp aniso}}$

$$\eta^2 \cdot \exp\left[-(\psi_1|w_1 - w_1'| + \psi_2|w_2 - w_2'|\right])$$

$\mathcal{GP}_{\text{additive}}$

$$\eta_1^2 \cdot \exp\left[-\psi_1|w_1 - w_1'|\right] + \eta_2^2 \cdot \exp\left[-\psi_2|w_2 - w_2'|\right]$$

$$+\eta_3^2 \cdot \exp\left[-(\psi_3|w_1 - w_1'| + \psi_4|w_2 - w_2'|)\right]$$

η^2 is measure of signal, w is warped input, ψ is bandwidth

Hyperpriors

$\mathcal{GP}_{\text{sq-exp iso}}$ & $\mathcal{GP}_{\text{sq-exp aniso}}$ Hyperpriors

$$\psi \sim \text{gamma}(4, 4)$$
$$\sigma^2 \sim \mathcal{T}^+(4, 0, 1)$$
$$\eta^2 \sim \mathcal{T}^+(4, 0, 1)$$

$\mathcal{GP}_{\text{additive}}$ Hyperpriors

$$\psi \sim \text{gamma}(4, 4)$$
$$\sigma^2 \sim \mathcal{T}^+(4, 0, 1)$$
$$\eta_i^2 | \lambda_i, \tau \sim \mathcal{N}^+(0, \lambda_i^2 \tau^2), \text{ for } i \in \{1, 2, 3\}$$
$$\lambda_i \sim \mathcal{T}^+(3, 0, 1)$$
$$\tau \sim \mathcal{C}^+(0, 1)$$

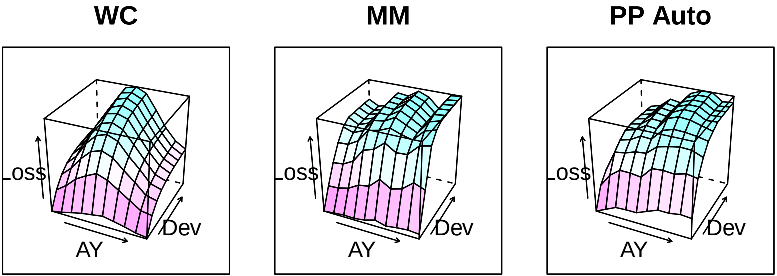
Model Goals Revisited

We would like to develop a model with

- Flexibility - relationship between accident and development years are determined by the data.
- Parsimony - Only a few hyperparameters for a very rich and robust model.
- Efficient computation - Can be implemented through standard software, we used STAN to take advantage of Hamiltonian Monte Carlo with automatic tuning from the No-U-Turn sampler. Code is available.
- Automation - Can be used without much tuning.

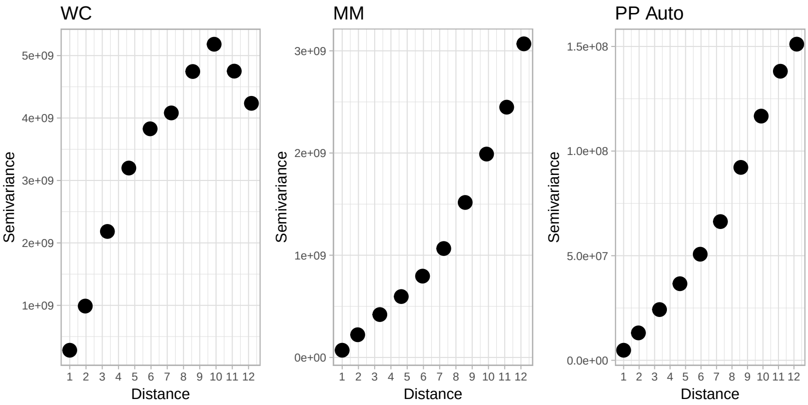
Data Analysis

Three Example Loss Triangles



State Farm Workers' Comp, Scpie Medical Malpractice, Farmers Private Passenger Auto

Variograms



WC
MM
PP Auto
Semivariance
Distance
5e+09
4e+09
3e+09
2e+09
1e+09
3e+09
2e+09
1e+09
0e+00
1.5e+08
1.0e+08
5.0e+07
0.0e+00
1 2 3 4 5 6 7 8 9 10 11 12

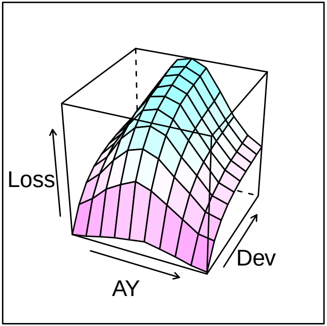
Point-wise Predictive Accuracy

	State Farm Workers' Comp	Scpie Indemnity Company Medical Malpractice	Farmers Automobile Group PP Auto
Chain Ladder	**7165**	15544	1685
Guszcza 2008 Weibull	11499	12753	1772
Guszcza 2008 Loglogistic	**9447**	15961	1940
$\mathcal{GP}_{\text{sq-exp iso}}$	13859	**6313**	**1401**
$\mathcal{GP}_{\text{sq-exp aniso}}$	10700	**6112**	**1436**
$\mathcal{GP}_{\text{additive}}$	**9478**	**6135**	**1390**

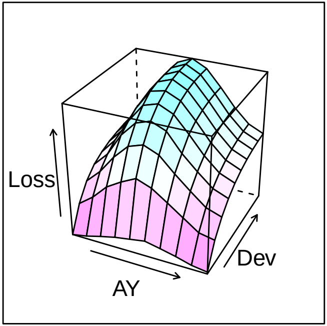
Comparing Predictive Accuracy in RMSE

Workers' Comp Predictions

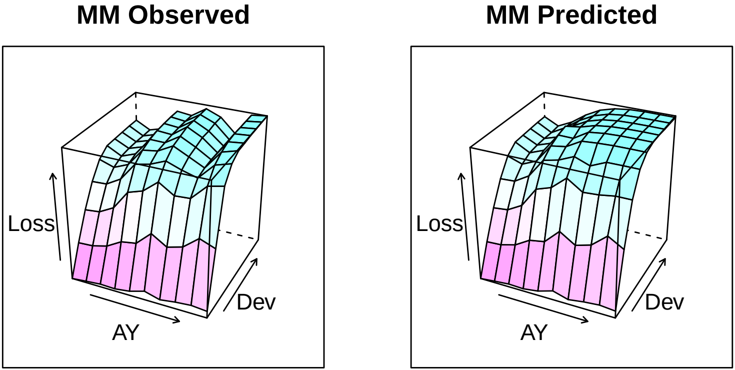
WC Observed



WC Predicted

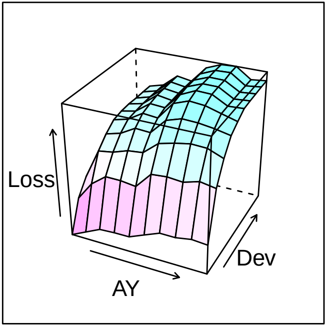


Medical Malpractice Predictions

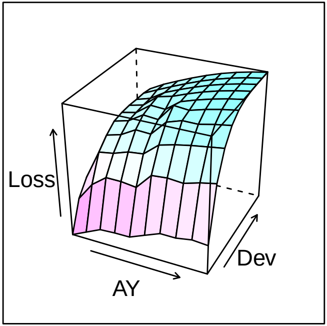


Private Passenger Auto Predictions

PP Observed



PP Predicted



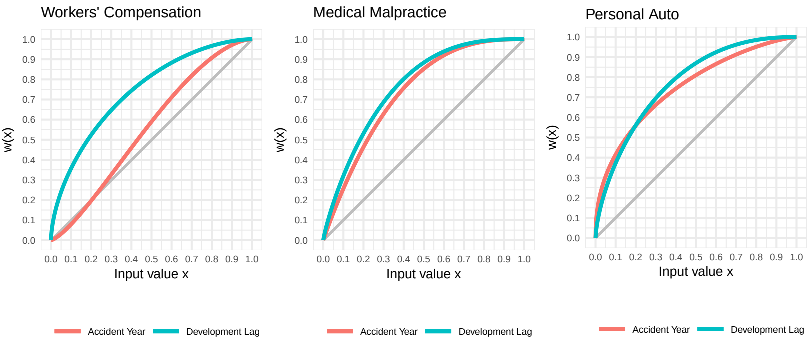
Ultimate Losses

	State Farm Workers' Comp		Scpie Indemnity Company Medical Malpractice		Farmers Automobile Group PP Auto	
	$\mathbb{E}$ [Ultimate Losses]	Difference	$\mathbb{E}$ [Ultimate Losses]	Difference	$\mathbb{E}$ [Ultimate Losses]	Difference
Chain Ladder	1614623	**-2928**	868122	73871	295962	5437
Guszcza 2008 Weibull	1546296	-71255	826759	32508	295852	5327
Guszcza 2008 Loglogistic	1618161	**610**	894115	99864	304347	13822
$\mathcal{GP}_{\text{sq-exp iso}}$	1635830	18279	774347	**-19904**	291034	**509**
$\mathcal{GP}_{\text{sq-exp aniso}}$	1621038	3487	780291	**-13960**	290687	**162**
$\mathcal{GP}_{\text{additive}}$	1617112	**-439**	794251	**-1919**	291625	**1100**
Observed	1617551		792332		290525	

Ultimate Loss Uncertainty

	State Farm Workers' Comp	Scpie Indemnity Company Medical Malpractice	Farmers Automobile Group PP Auto
Observed 10 Year Ultimate Losses	1617551	792332	290525
$\mathcal{GP}_{\text{sq-exp iso}}$ Predictive Interval	(1374281, 1897872), 523591	(674193, 850110), 175917	(277161, 304586), **27425**
$\mathcal{GP}_{\text{sq-exp aniso}}$ Predictive Interval	(1419957, 1791636), **371679**	(698744, 856794), 158050	(274613, 308932), 34319
$\mathcal{GP}_{\text{additive}}$ Predictive Interval	(1402950,1846652), 443702	(722815, 858862), **136047**	(275758, 304859), 29101

Input Warping



Estimating Loss Reserves Using Hierarchical Bayesian Gaussian Process Regression with Input Warping

Brian Hartman (Brigham Young University)

Joint work with
Nathan Lally (Pratt and Whitney)

Advances in Predictive Analytics
01 December 2017

Paper and code available at `hartman.byu.edu`