Gaussian Processes for Modeling Loss Development

UCSB InsurTech Summit

Howard Zail, Partner, Elucidor, LLC

May 3, 2019

Outline

- The Challenge: Extrapolating Paid Loss Triangles
- Introduction to Gaussian Processes
- The Basic GP Setup for Loss Modeling
- Advanced Models
- Performance Metrics
- Multiple Company Models

Based on a paper written with Prof. Ludkovski

The Challenge

Completing the Square: A Fundamental Process

Upper Left Triangle: Training Data

Accident Year	Development Lag									
	1	2	3	4	5	6	7	8	9	10
1988	952	1,529	2,813	3,647	3,724	3,832	3,899	3,907	3,911	3,912
1989	849	1,564	2,202	2,432	2,468	2,487	2,513	2,526	2,531	
1990	983	2,211	2,830	3,832	4,039	4,065	4,102	4,155		
1991	1,657	2,685	3,169	3,600	3,900	4,320	4,332			
1992	932	1,940	2,626	3,332	3,368	3,491				
1993	1,162	2,402	2,799	2,996	3,034					
1994	1,478	2,980	3,945	4,714						
1995	1,240	2,080	2,607							
1996	1,326	2,412								
1997	1,413									

Completing the Square: A Fundamental Process

Lower Right Triangle: Extrapolation

Accident Year	Development Lag									
	1	2	3	4	5	6	7	8	9	10
1988										
1989										2,527
1990									4,268	4,274
1991								4,338	4,341	4,341
1992							3,531	3,540	3,540	3,583
1993						3,042	3,230	3,238	3,241	3,268
1994					5,462	5,680	5,682	5,683	5,684	5,684
1995				3,080	3,678	4,116	4,117	4,125	4,128	4,128
1996			3,367	3,843	3,965	4,127	4,133	4,141	4,142	4,144
1997		2,683	3,173	3,674	3,805	4,005	4,020	4,095	4,132	4,139

Completing the Square: A Fundamental Process

Completed Square

Accident Year	Development Lag									
	1	2	3	4	5	6	7	8	9	10
1988	952	1,529	2,813	3,647	3,724	3,832	3,899	3,907	3,911	3,912
1989	849	1,564	2,202	2,432	2,468	2,487	2,513	2,526	2,531	2,527
1990	983	2,211	2,830	3,832	4,039	4,065	4,102	4,155	4,268	4,274
1991	1,657	2,685	3,169	3,600	3,900	4,320	4,332	4,338	4,341	4,341
1992	932	1,940	2,626	3,332	3,368	3,491	3,531	3,540	3,540	3,583
1993	1,162	2,402	2,799	2,996	3,034	3,042	3,230	3,238	3,241	3,268
1994	1,478	2,980	3,945	4,714	5,462	5,680	5,682	5,683	5,684	5,684
1995	1,240	2,080	2,607	3,080	3,678	4,116	4,117	4,125	4,128	4,128
1996	1,326	2,412	3,367	3,843	3,965	4,127	4,133	4,141	4,142	4,144
1997	1,413	2,683	3,173	3,674	3,805	4,005	4,020	4,095	4,132	4,139

Key Modeling Goals

- Determine a reserve for unpaid claims
  - (Sum of Column 10 completed square) - (paid claims)
- Distribution of reserves
  - For allocating risk capital
- Expected value of successive step-ahead paid claims
  - For cash flow projections & asset / liability management

Current Modeling Landscape

- LDF Models
  - Ubiquitous, most popular, straightforward
  - Chain Ladder and variants
  - Good at producing expected claims
  - Underestimates uncertainty around claims
- ILR models
  - Overdispersed Poisson Model
  - Can be done using standard GLM software

Introduction to Gaussian Processes

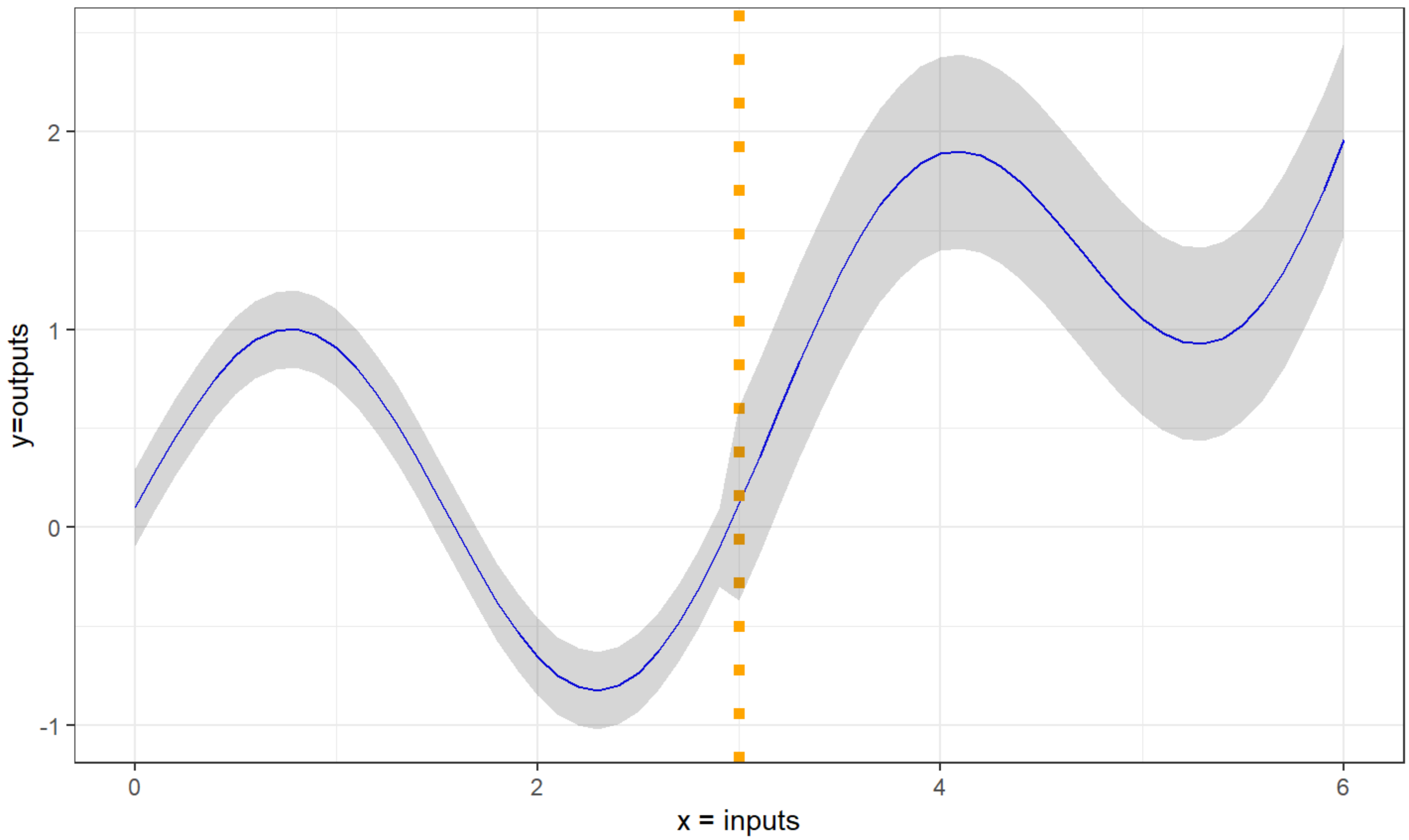
What is a Gaussian Process

- Multivariate Normal distribution:
  - Mean vector, covariance matrix
- Gaussian Process:
  - Finite sample is Multivariate Normal
  - Mean function, covariance function

Modeling Thought Process with GPs

Data with inputs that are "near" to one another, will have outputs that are "near".

A Toy Example



Model Specification

Ground Truth

- x inputs, Y output
- True $Y(x) \sim N(\mu(x), \sigma(x))$
- $\mu(x) = 0.1(x-1)^2 + sin(2x)$
- $\sigma(x) = (.1).1_{\{x<3\}} + (.25).1_{\{x\geq 3\}}$

Model

- $x = vector,\ x^i = i^{th}$ element of x
- $f(x) \sim GP(m(x), C(x, x'))$
- $\text{Modeled } Y(x) = f(x) + \epsilon,\ \text{where } \epsilon \sim N(0, \sigma)$

A Toy Example: "Nearness" defined in covariance function

$$C = \text{Covariance Matrix}$$
$$C_{i,j} = i, j^{th} \text{element of C}$$
$$C_{i,j} = \eta^2 . e^{-(x^i - x^j)^2/\rho^2}$$

- When x^i is close to x^j, y^i will have a high covariance with y^j.
- When x^i is far from x^j, y^i will have a low covariance with y^j.
- C is called a Squared Exponential Kernel
- η is the amplitude
- ρ is the lengthscale

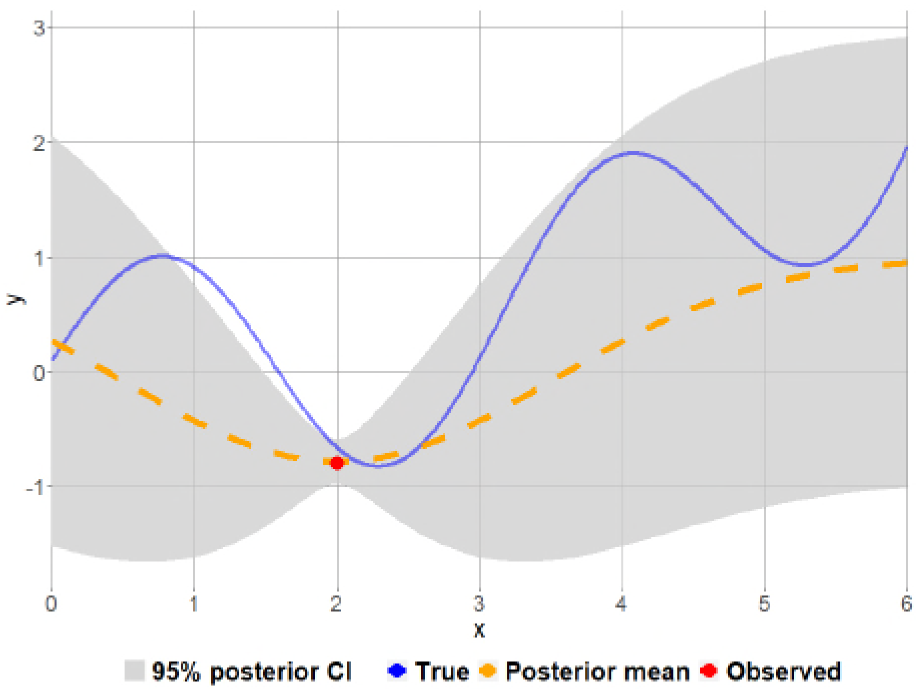
An Advantage of GPs

- If the prior $f(x) \sim GP(m(x), C(x, x))$, then
- Posterior $f_*(x_*)|D \sim GP(m*(x_*), C*(x_*, x_*))$, where closed form expressions can be obtained for m_* and C_*

A Toy Example: Choose a set of Parameters

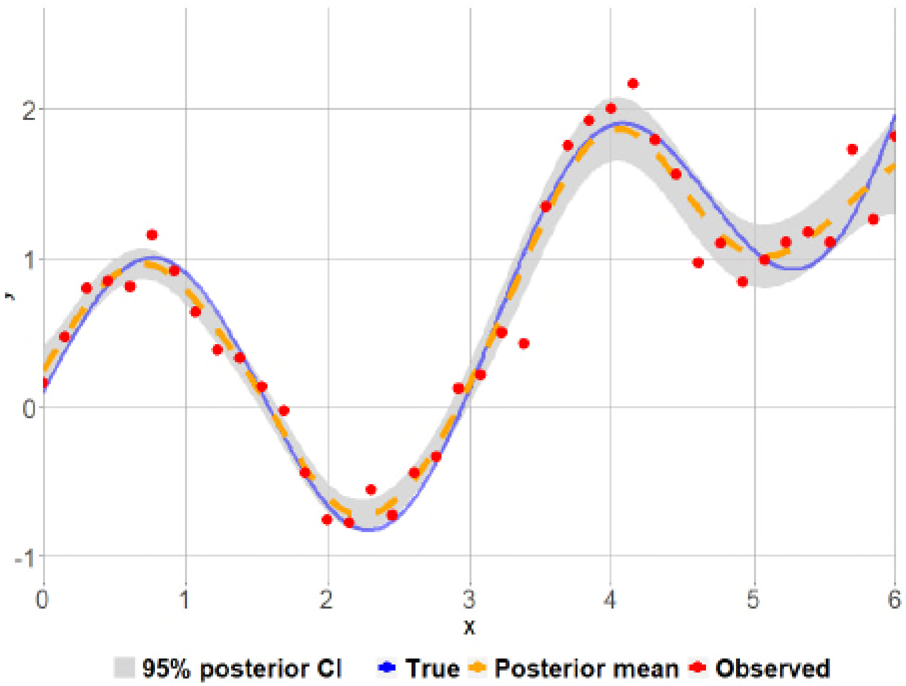
$$
\begin{aligned}
m(x) &= 1 \\
\eta &= 1 \\
\rho^2 &= 1.5 \\
\sigma &= 1
\end{aligned}
$$

A Toy Example: One Training Sample



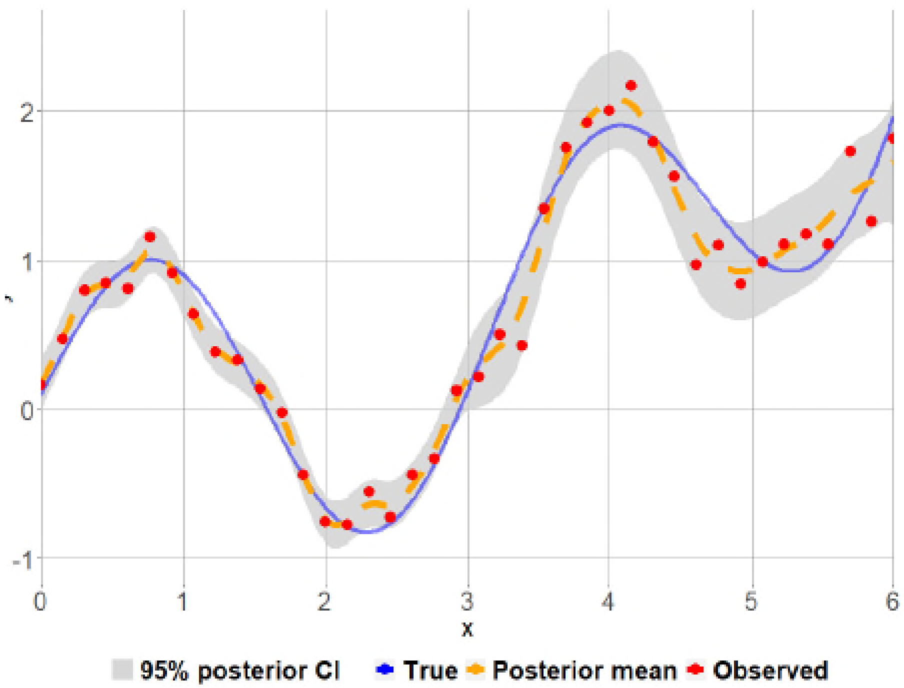
- Blue line: True mean; Red Dots: Samples
- Dotted Orange: Posterior GP; Grey Band: 95% credible interval

A Toy Example: 40 Training Samples



A Toy Example: 40 Training Sample, with shorter lengthscale

Use $\rho^2 = 0.5$ instead of $\rho^2 = 1.5$



The Basic GP Setup

GPs with Loss Reserves

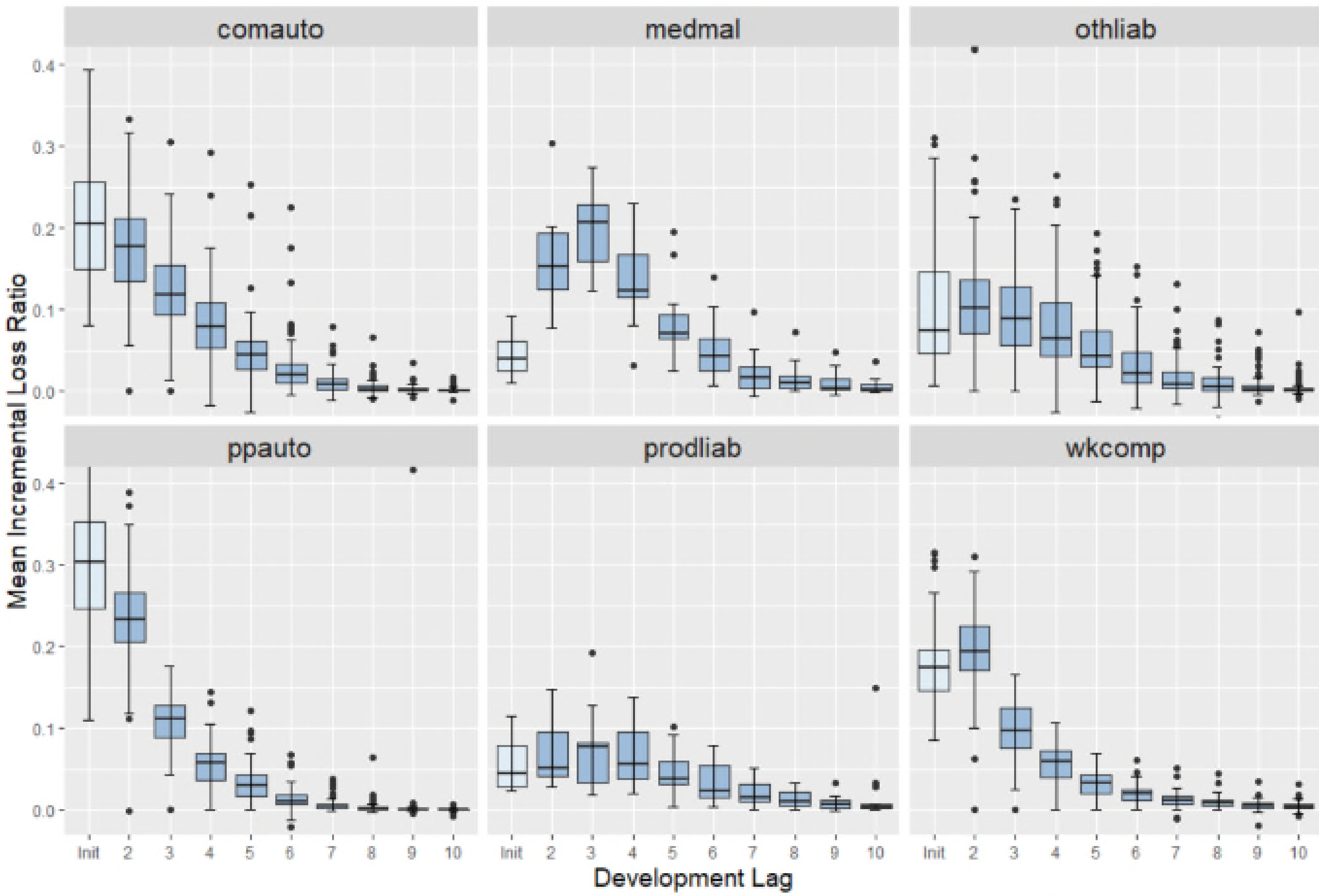
Each "cell" in a triangle is treated as a data point:

- $x^i = (AY^i, DL^i)$
- $L^i = \text{incremental loss ratio}$
- $L^i = f(x^i) + \epsilon_q$, where $q = \text{DL}$
- Prior $f(x^i) \sim GP(m(x^i), C(x, x))$
- $\epsilon_q \sim N(0, \sigma_q^2)$

The Data We Analyzed

- NAIC Schedule P triangles (Meyers & Shi):
  - We use *paid* claims only
- Completed squares from 1998-2006
- 10 accident years, 10 development years
- 200 companies across 6 business lines
- comauto (84), medmal (12), ppauto (96), prodliab (87), other (13), wkcomp (57)
- Training data (upper left triangle):
  - 55 ILR cells
- Test data (lower right triangle):
  - 45 ILR cells

Distribution of ILRs



Complexity in the Data

- ILRs are generally declining monotonically by Development Lag (with some exceptions)
- ILRs/paid claims are almost always positive
- Intrinsic uncertainty ($f(x)$) declines by Development Lag
- Extrinsic uncertainty (σ) declines by Development Lag
- Uncertainty is skewed to the right

The GP Models

Choice of Software

- Used Stan as our probabilistic programming framework
- Full Bayesian implementation
- Uses HMC as its core algorithm
- Solid, tested, strong online community
- Compared to DiceKriging package, but Stan offers far more flexibility
- Tried Greta:
    - Based on Tensorflow Probability.
    - Utilizes GPUs, a significant advantage
    - Still too early in production, so it was abandoned

Compound Kernels

- GPs can be used to do Bayesian linear regression by using a special kernel
- With Bayesian Regression we specify a mean function:
  - Correlation among the data points is an byproduct of the model
- With GP regression, we can specify a kernel
  - mean is a byproduct of the model e.g.
- Bayesian Linear Regression:
  - $Y(x) = ax + b$
  - $a \sim N(0, \sigma_a^2)$
  - $b \sim N(0, \sigma_b^2)$
- This is equivalent to:
  - $Y(x) \sim GP(0, \sigma_a.(x.x') + \sigma_b^2)$
  - Called "Linear Kernel"
- Kernels can be combined to form a compound kernel

"Plain" ILR GP Model

$$
\begin{aligned}
L(x) &= \text{Incremental Loss Ratio} \\
L(x) &= f(x) + \epsilon_q \\
f(x) &\sim GP(m(x), C(x,x)) \\
m(x) &= 0 \\
C(x,x) &= SquareExp(AY, DL) + \\
&\quad Linear(AY) + \\
&\quad Linear(log(DL)) \\
\epsilon_q &\sim N(0, \sigma_q) \\
\sigma_q &\geq \sigma_{q+1}
\end{aligned}
$$

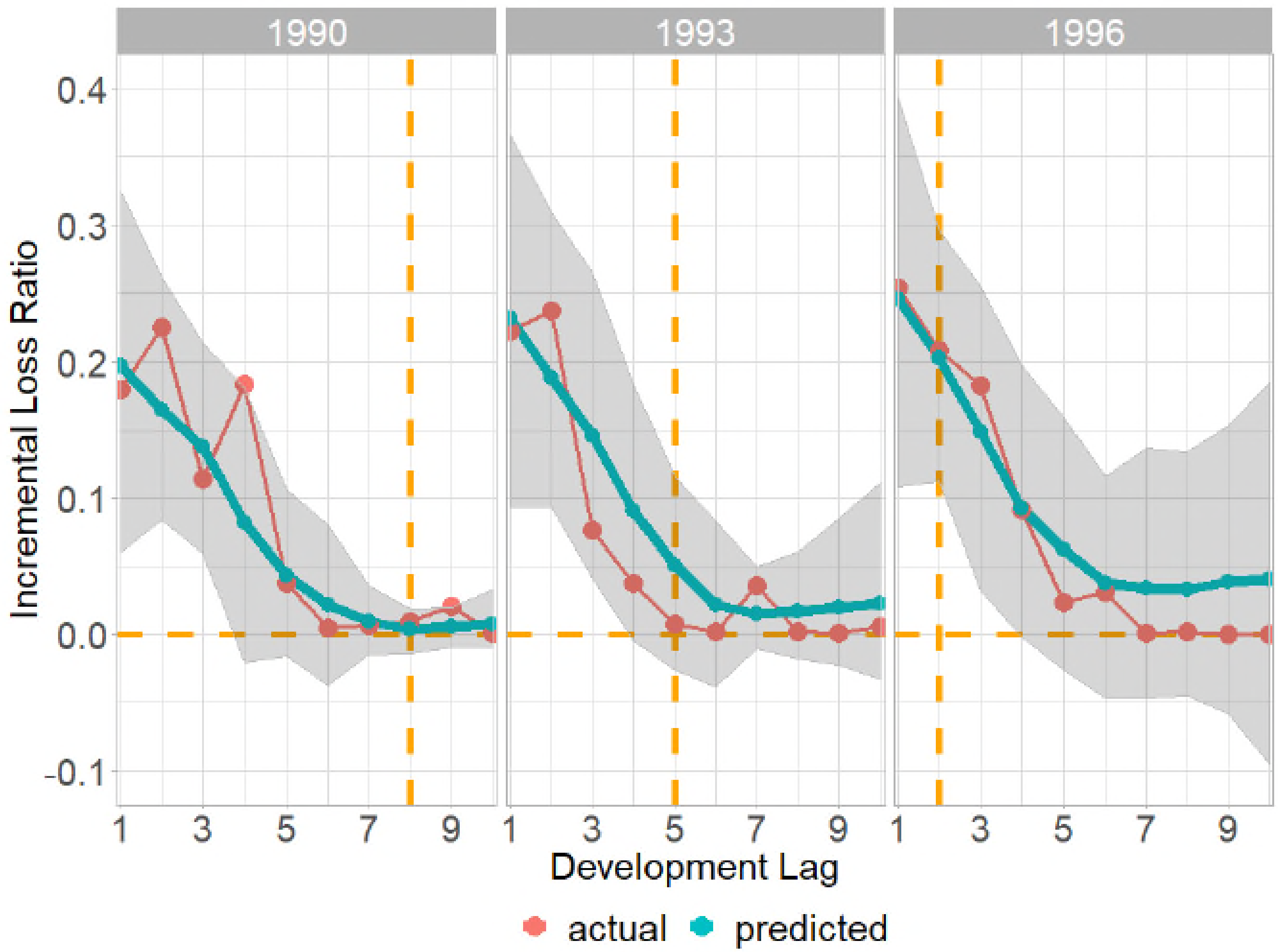
- Both input and output data were standardized

Source of Uncertainty: The Three Sources

1. Extrinsic Uncertainty:
   - ϵ_q
2. Intrinsic Uncertainty:
   - Uncertainty in $f(x)$, modeled as a Gaussian random field
3. Correlation Uncertainty (model misspecification):
   - Uncertainty in the hyperparameters (ρ, η, etc.)

Overall predictive uncertainty no longer under-estimated

Sample fit of "Plain" ILR GP Model



Problems with Plain ILR GP

- Large negative paid losses
- **Increasing** uncertainty for large Development Lags, but we know from data that there is *decreasing* uncertainty at the long Development Lags.
- Mean prediction is not declining asymptotically to zero

Advanced Models

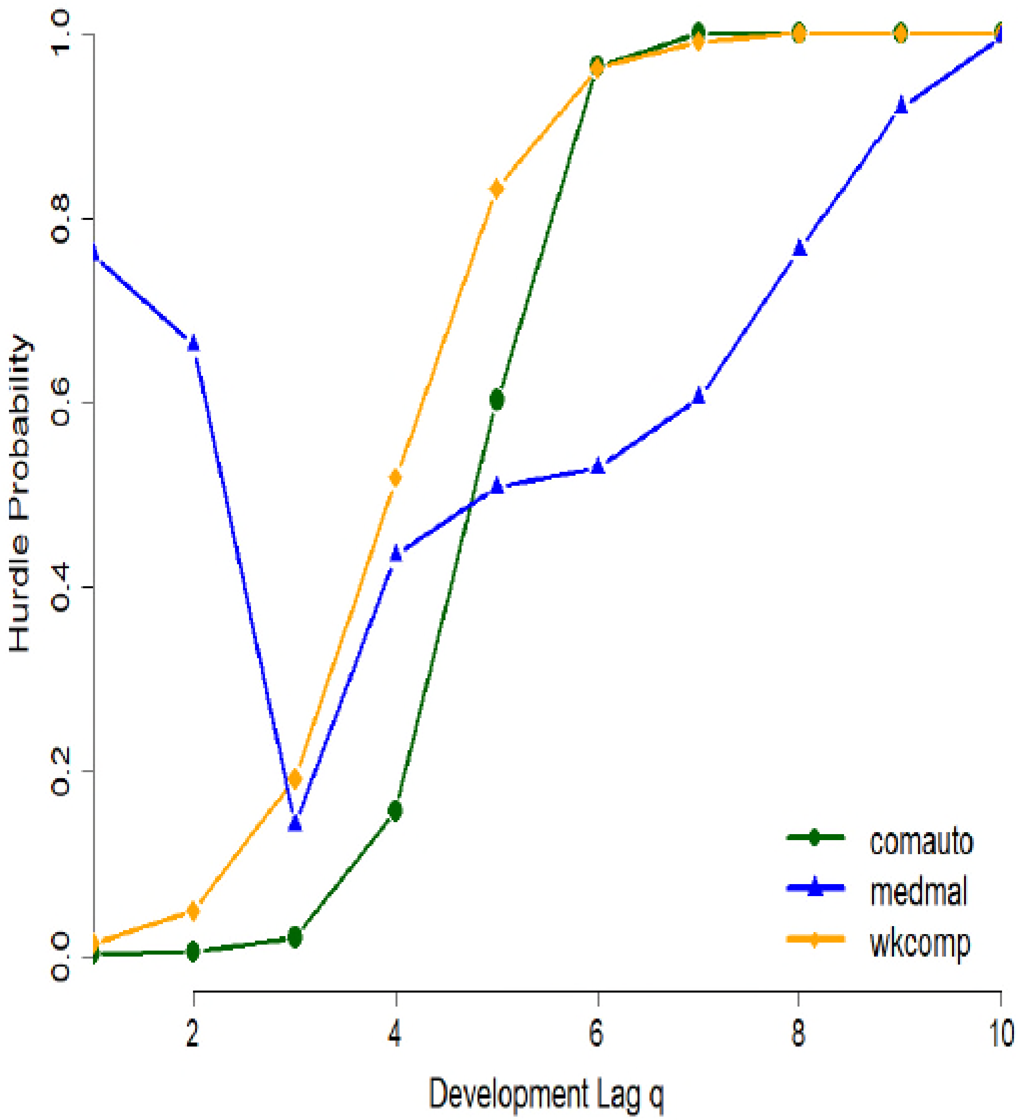
Enhancement 1: Hurdle Model

$$L_{hurdle}(x) = max(0, L(x))$$

- Mixture of a truncated normal and a point mass at zero.
- The respective likelihood $L_{hurdle}(x)|f(x)$ during the inference step is a mixture of:
  - a Gaussian likelihood if $L(x) > 0$ and
  - a Bernoulli likelihood:

$P(L_{hurdle}(x) = 0|f(x)) = \Phi(0; mean = f(x), var = \sigma_q^2(x))$, called the hurdle probability.

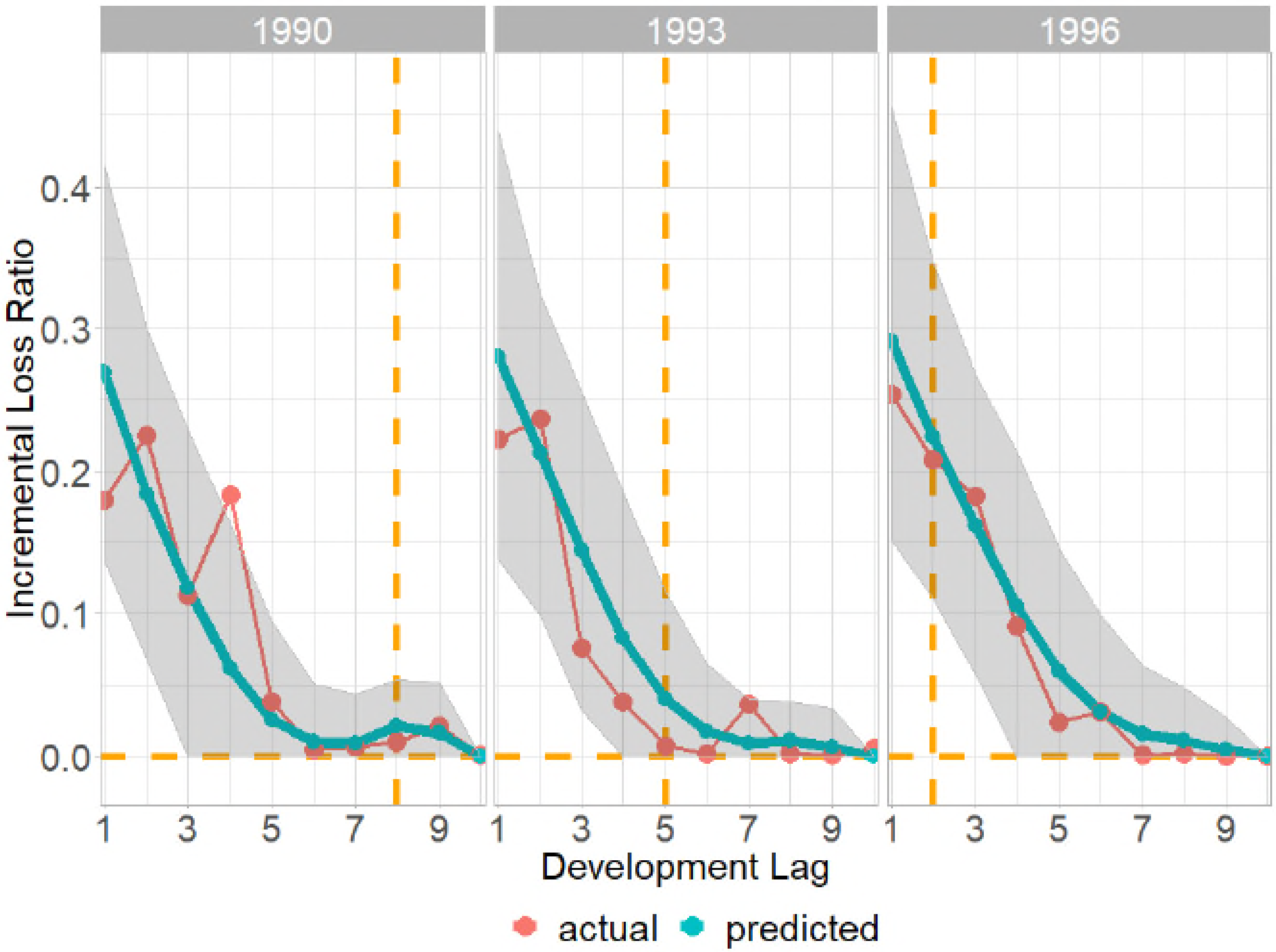
Hurdle Probabilities



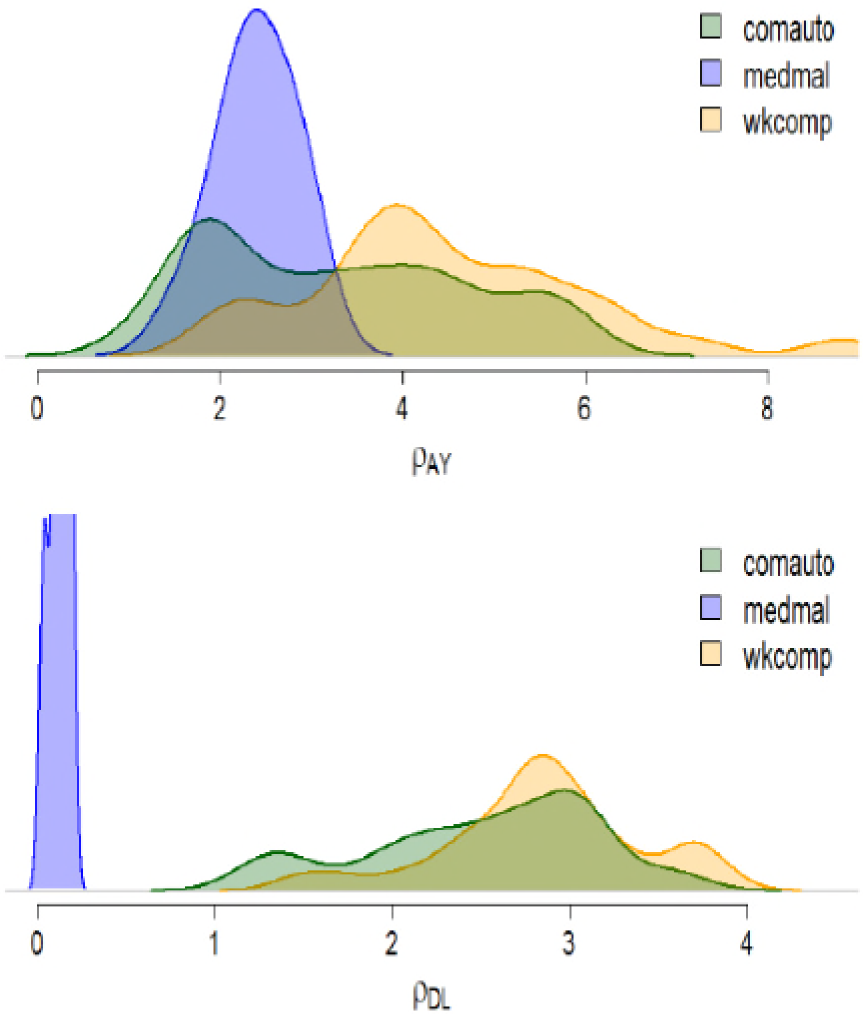
Enhancement 2: Virtual Datapoints

- At Development Lag q = 11,we add virtual observations of zero ILRs
- Equivalent to telling model that losses have fully developed by DL=10
- This partly turns an extrapolation model into an interpolation model
- (In practice, an actuary can use non-zero virtual observations for longer tailed business lines)

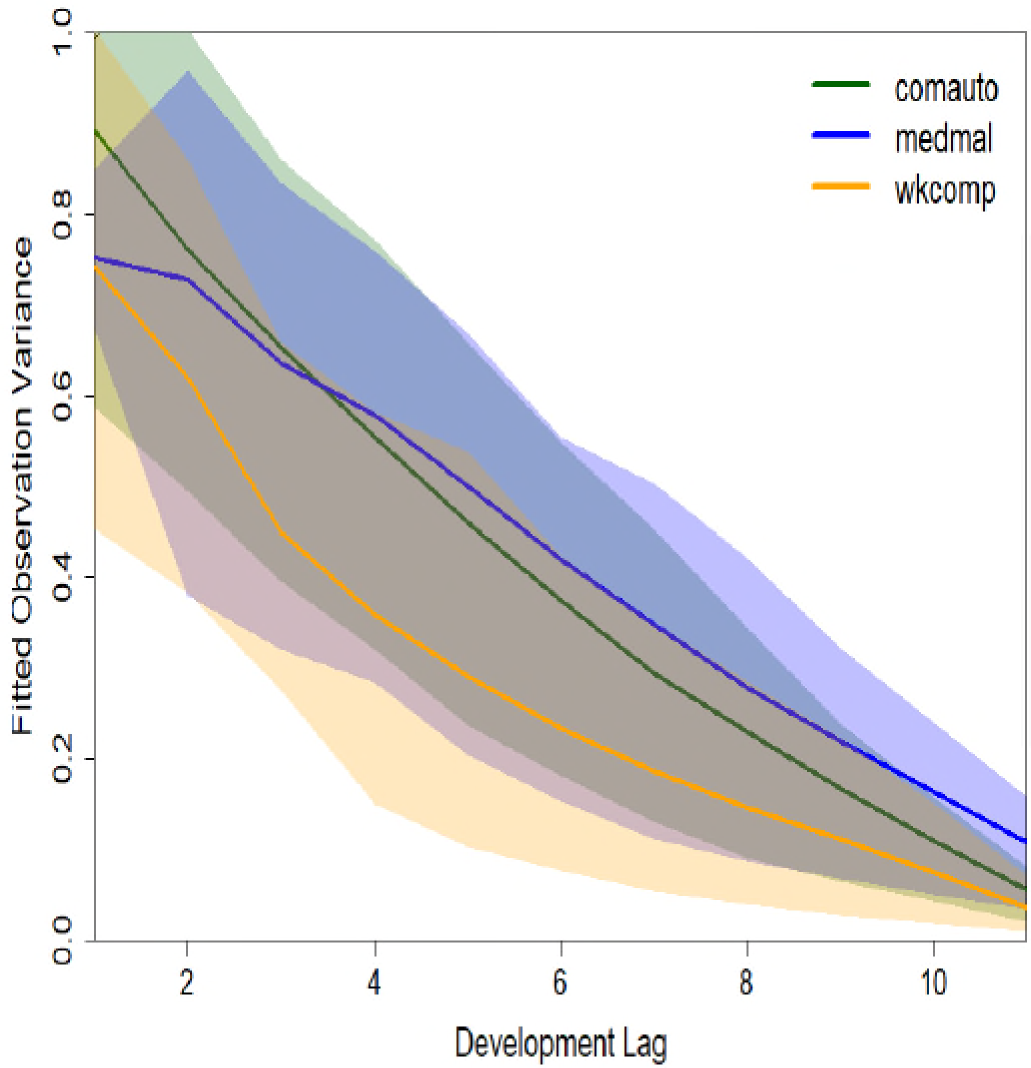
Sample fit of Hurdle+Virutal ILR GP Model



Lengthscales under the GP Model

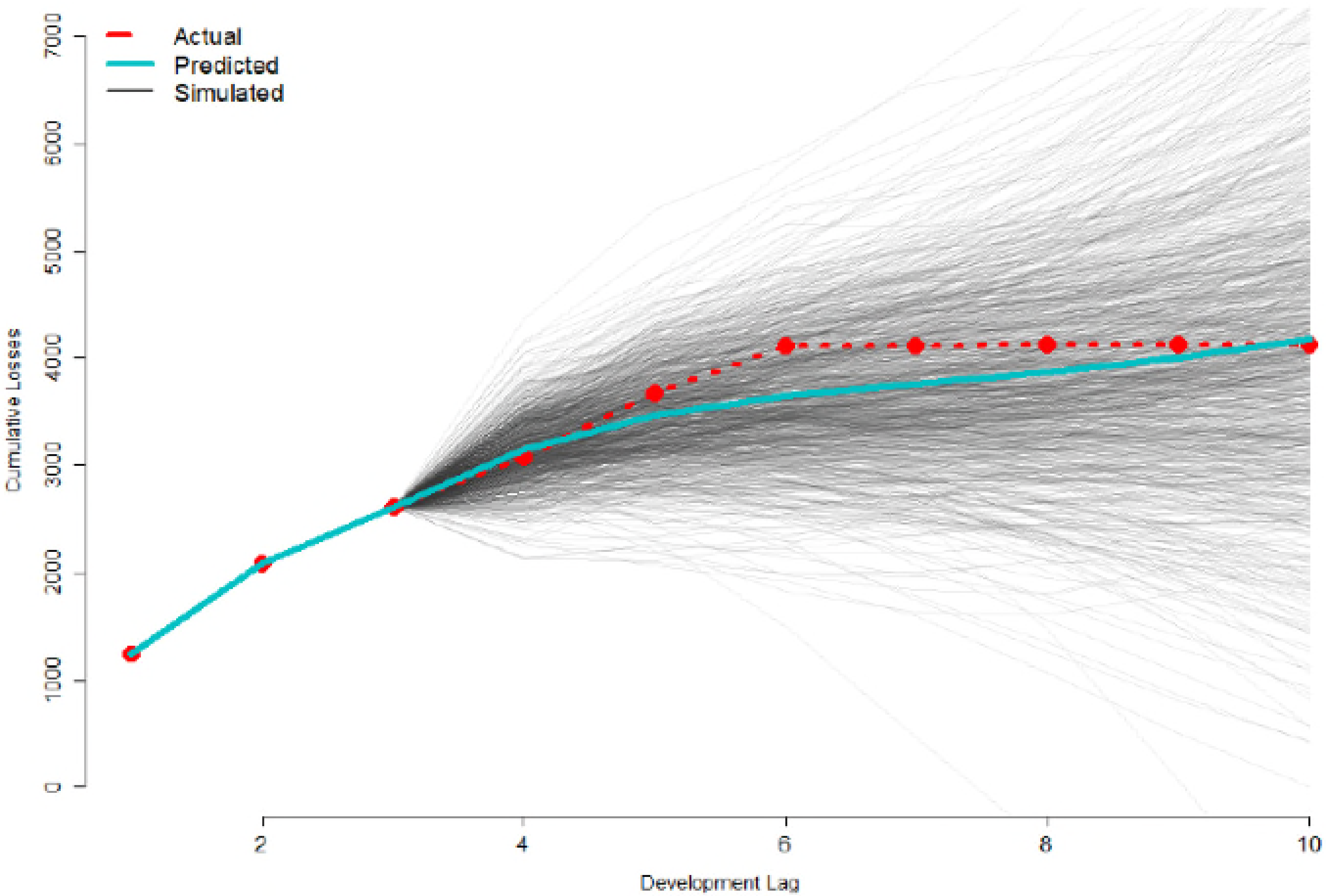


Variance under ILR GP Model



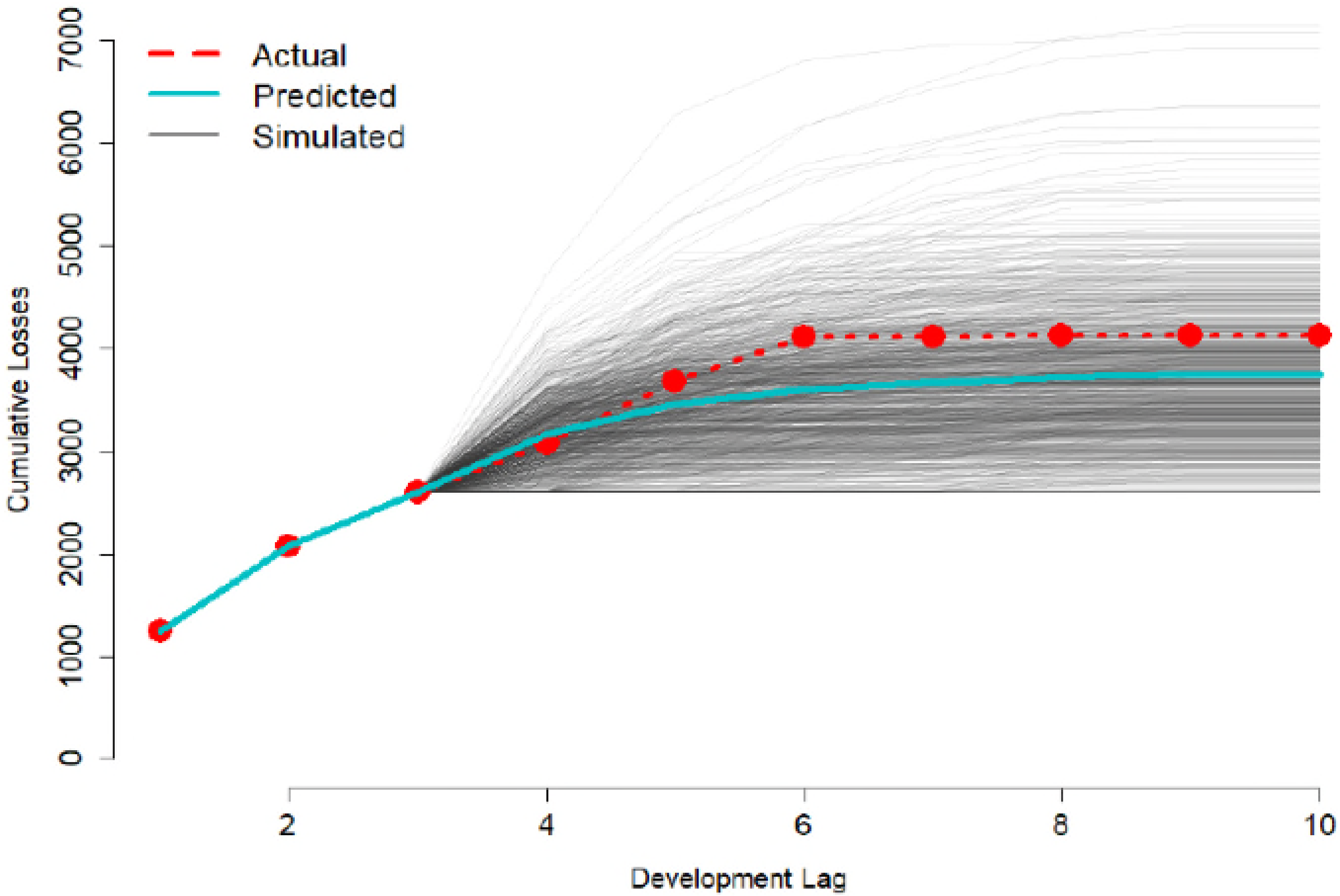
Comparison of Plain vs. Hurdle+Virutal GP

Simulations from the Plain GP Model

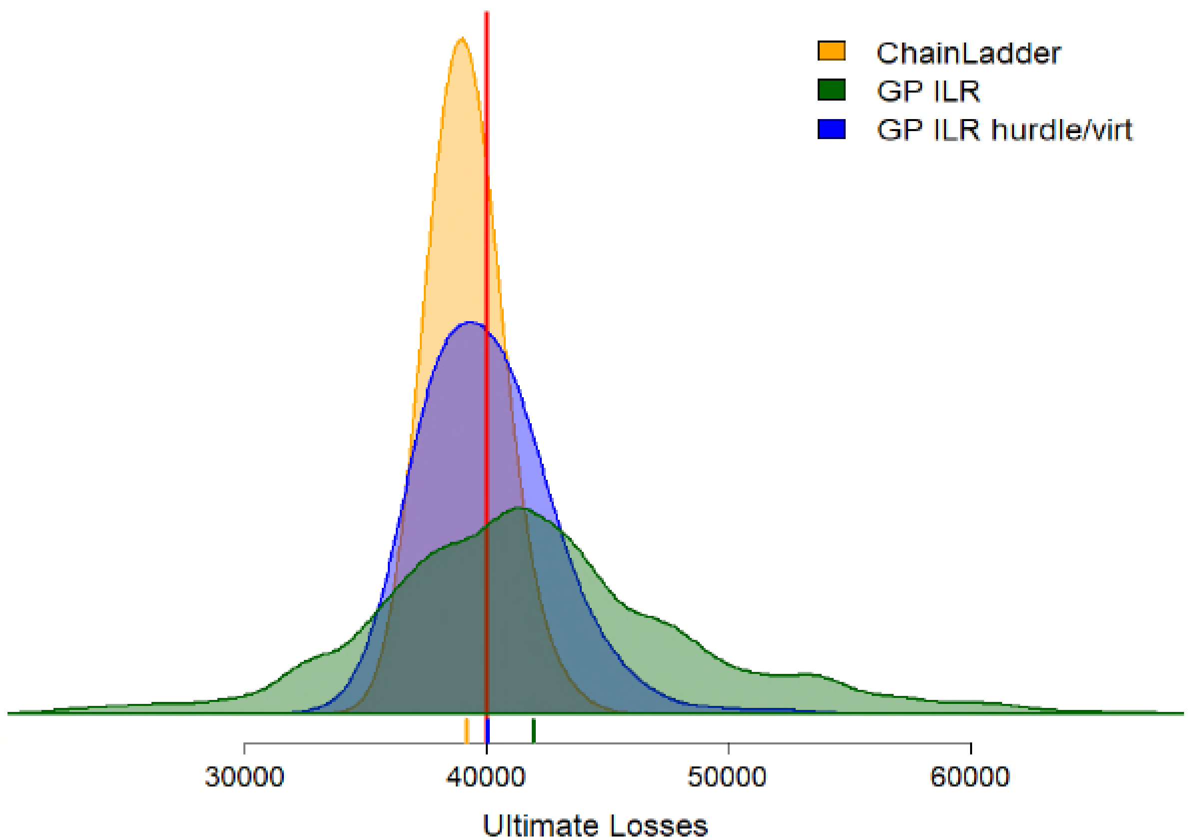


Comparison of Plain vs. Hurdle+Virutal GP

Simulations from the Hurdle + Virtual Data GP Model



Comparison of Chain Ladder vs. Plain vs. Hurdle+Virutal GP



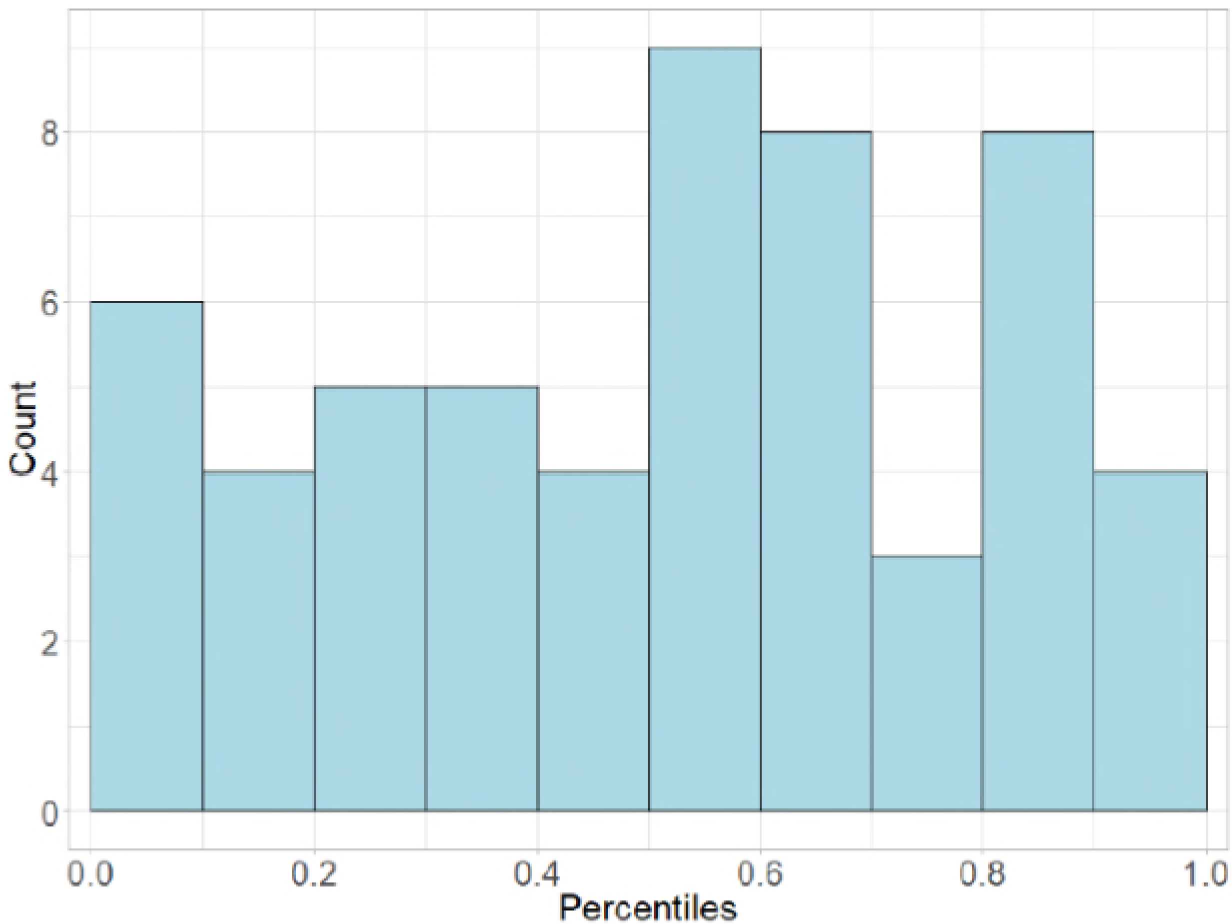
Performance Metrics

- Best estimate reserves: RMSE of ultimate loss ratio
- Cash flow projections: RMSE of step-ahead cumulative loss ratio
- Risk Capital: Coverage Ratio
- Sensitivity Analysis:
  - CRPS (Continuous Ranked Probability Score)
  - NLPD (negative log probability density)
- Kolmogorov-Smirnov Test
- Graphical Tests (distribution of percentile ranks)

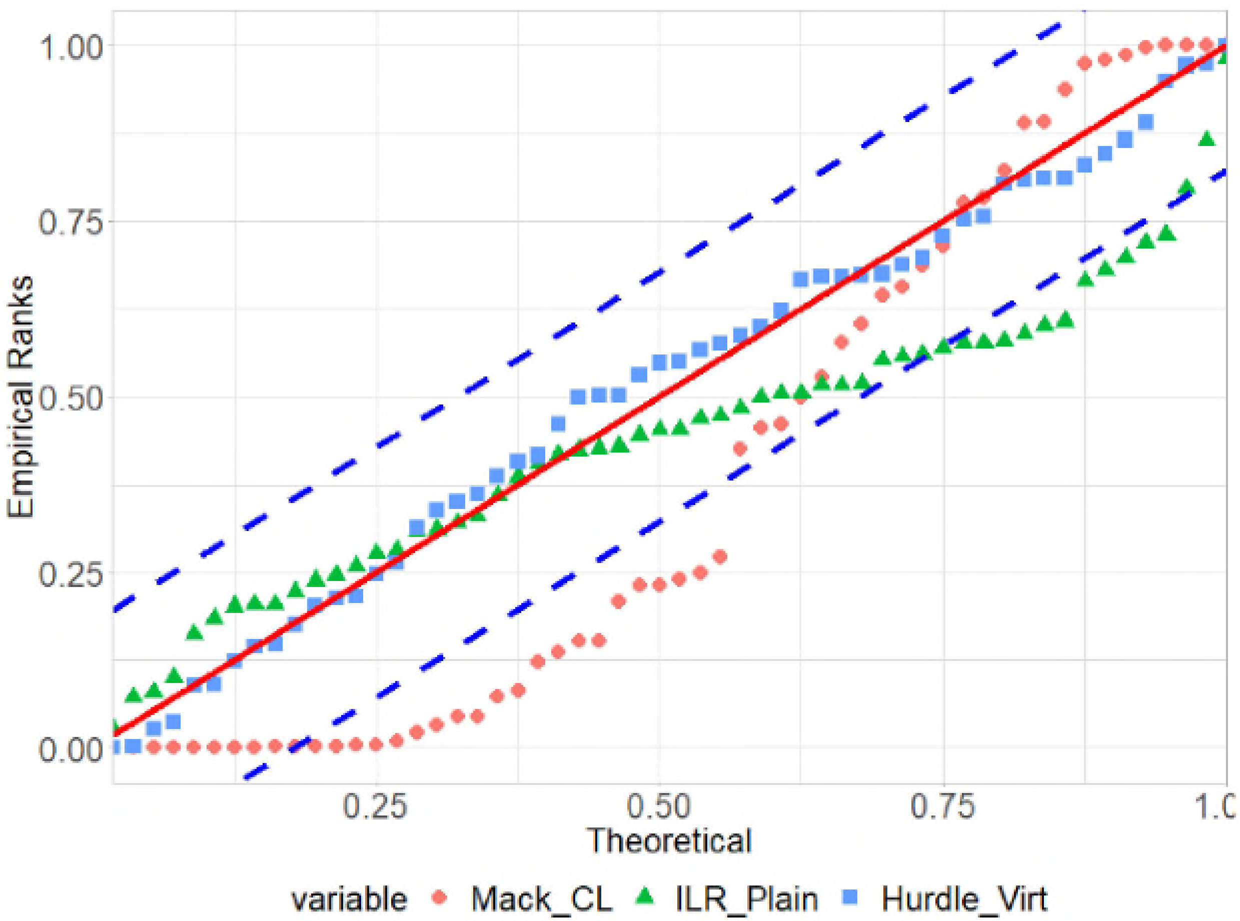
Sample Test: WKCOMP RMSE and Coverage

Algorithm	Total RMSE	LR RMSE	Coverage
Mack CL	24726.2	0.049	0.509
Bootstrap CL	24895.7	0.052	0.544
ILR Plain	79812.0	0.138	0.965
ILR Hurdle	41220.3	0.115	0.930
ILR Hurdle+Virt	42096.4	0.088	0.875

Sample Test: WKCOMP Distribution of Percentiles



Sample Test: WKCOMP Kolmogorov Smirnov Test



Multiple Company Models

- GP Model can be extended to handle multiple companies within the same model
- Companies can "borrow strength" from other companies
- Find high correlation in ILRs among companies
- Difficult (computationally) to model many companies, but a reasonable set of comparable companies can be added into a single model

The Multiple Company Model

$$C^{multi}(x^i, x^j) = C^{single}(x^i, x^j).\, e^{-\rho_{co}.(1-\delta_{i,j})},\ \text{where}$$

$$\delta_{i,j} = \begin{cases} 0, & Company^i \neq Company^j \\ 1, & \text{otherwise} \end{cases}$$

The Advanced Multiple Company Model

$$C^{multi}(x^i, x^j) = C^{single}(x^i, x^j).\, e^{-(\rho_{co,i}+\rho_{co,j}).(1-\delta_{i,j})},\ \text{where}$$

$$\delta_{i,j} = \begin{cases} 0, & Company^i \neq Company^j \\ 1, & \text{otherwise} \end{cases}$$

Multiple Company Model Scores

$$\rho^T_{Co,\cdot} = \begin{bmatrix} 0.0175 \\ 0.02810 \\ 0.1347 \\ 0.2744 \\ 0.03394 \end{bmatrix}, \quad R^{(Co)} := \begin{pmatrix} 1 & & & & \\ 0.9553 & 1 & & & \\ 0.8587 & 0.8497 & 1 & & \\ 0.7467 & 0.7389 & 0.6641 & 1 & \\ 0.9498 & 0.9398 & 0.8449 & 0.7346 & 1 \end{pmatrix}$$

Overall Model Comparisons

- RMSE under Mack CL beat Hurdle+Virtual model 4 out 6 business lines
- Coverage under Hurdle+Virtual model beat CL 5 out 6 business lines

Summary

We introduced a framework for modeling complexities of paid loss development

- Captures extrinsic, intrinsic and correlation risk
- Comply with structural data constraints and complexities
- Cohently project full multi-period trajectory of losses
- Borrow strength from other company data in a multi-company model
- Provided a comprehensive means of testing model fit

Overall GPs can provide a material improvement in understanding the risk profile of liabilities