Experience Rating in Insurance Pricing

Mario V. Wüthrich
RiskLab, ETH Zurich

Joint work with Jaeyoun Ahn, Himchan Jeong, Yang Lu

Insurance Data Science Conference

Stockholm University

June 17-18, 2024

Overview

There is an increasing interest in further developing experience rating.

Table of contents:

- Prior information rating
- Posterior rating: static case
- Posterior rating: dynamic case
- Deep experience rating

- **Section 1: Prior information rating**

Best-estimate (actuarial fair) pricing

- **Aim:** Price an insurance claim Y based on prior rating information $\boldsymbol{x}$.
- Prior rating information $\boldsymbol{x}$ is available at the inception of the insurance contract, e.g., age of policyholder, place of living, price of insured object, etc.
- Prior rating information is also called covariates or (static) features.
- Best-estimate price for claim Y, given prior rating information $\boldsymbol{x}$,

$$\boldsymbol{x} \;\mapsto\; \mu(\boldsymbol{x}) = \mathbb{E}\left[Y \middle| \boldsymbol{x}\right].$$

- **Actuarial task:** Estimate this pricing functional (using past data).
- Popular approaches: generalized linear models (GLMs) or neural networks.

Best-estimate (actuarial fair) pricing

- **Aim:** Price an insurance claim Y based on prior rating information $\boldsymbol{x}$.

- Prior rating information $\boldsymbol{x}$ is available at the inception of the insurance contract, e.g., age of policyholder, place of living, price of insured object, etc.

- Prior rating information is also called covariates or (static) features.

- Best-estimate price for claim Y, given prior rating information $\boldsymbol{x}$,

$$\boldsymbol{x} \;\mapsto\; \mu(\boldsymbol{x}) = \mathbb{E}\left[Y \middle| \boldsymbol{x}\right].$$

- **Actuarial task:** Estimate this pricing functional (using past data).

- Popular approaches: generalized linear models (GLMs) or neural networks.

- **Section 2: Posterior rating: static case**

Experience rating

- What if past claims history $Y_{1:t} = (Y_1, \ldots, Y_t)$ is available to predict Y_{t+1}?

- Posterior/experience rating considers

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}, \boldsymbol{x}_{1:t+1}\right],$$

  or if no prior rating information is available

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}\right].$$

- This is also known as random effects modeling.

- Such models are useful if there is dependence between Y_{t+1} and $Y_{1:t}$.

- This dependence can be of a static or of a dynamic nature.

Random effects: static case

- The most popular experience rating models belong to the exponential dispersion family (EDF) with conjugate priors; Bichsel (1964), Jewell (1974).

- The Bühlmann–Straub (BS) model (1970) gives a linear (credibility) approximation in case of intractable posterior distributions.

- The BS model essentially assumes for all time periods $1 \leq s \leq t+1$

$$\mathbb{E}\left[\left.Y_s\right|\Theta\right] \;=\; \mu(\Theta),$$

with common latent (risk) factor Θ (+ conditional independence assumptions).

- This is the static case as the latent factor Θ does not dependent on time s.

- For experience rating we need to compute Bayes' formula

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[\left.Y_{t+1}\right|Y_{1:t}\right] = \mathbb{E}\left[\left.\mu(\Theta)\right|Y_{1:t}\right].$$

Random effects: static case

- The most popular experience rating models belong to the exponential dispersion family (EDF) with conjugate priors; Bichsel (1964), Jewell (1974).

- The Bühlmann–Straub (BS) model (1970) gives a linear (credibility) approximation in case of intractable posterior distributions.

- The BS model essentially assumes for all time periods $1 \leq s \leq t+1$

$$\mathbb{E}\left[Y_s \middle| \Theta\right] = \mu(\Theta),$$

with common latent (risk) factor Θ ($+$ conditional independence assumptions).

- This is the static case as the latent factor Θ does not dependent on time s.

- For experience rating we need to compute Bayes' formula

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}\right] = \mathbb{E}\left[\mu(\Theta) \middle| Y_{1:t}\right].$$

Random effects: static case

- The most popular experience rating models belong to the exponential dispersion family (EDF) with conjugate priors; Bichsel (1964), Jewell (1974).

- The Bühlmann–Straub (BS) model (1970) gives a linear (credibility) approximation in case of intractable posterior distributions.

- The BS model essentially assumes for all time periods $1 \leq s \leq t+1$

$$\mathbb{E}\left[\left. Y_s \right| \Theta\right] = \mu(\Theta),$$

  with common latent (risk) factor Θ ($+$ conditional independence assumptions).

- This is the static case as the latent factor Θ does not dependent on time s.

- For experience rating we need to compute Bayes' formula

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[\left. Y_{t+1} \right| Y_{1:t}\right] = \mathbb{E}\left[\left. \mu(\Theta) \right| Y_{1:t}\right].$$

Bühlmann–Straub credibility estimator

- The BS credibility estimator is given by

$$\widehat{\mu}^{\text{post}}_{Y_{t+1}|Y_{1:t}} \;=\; \omega_t \bar{Y}_t + (1-\omega_t)\,\mu_0,$$

with credibility weights and observation based estimators, respectively,

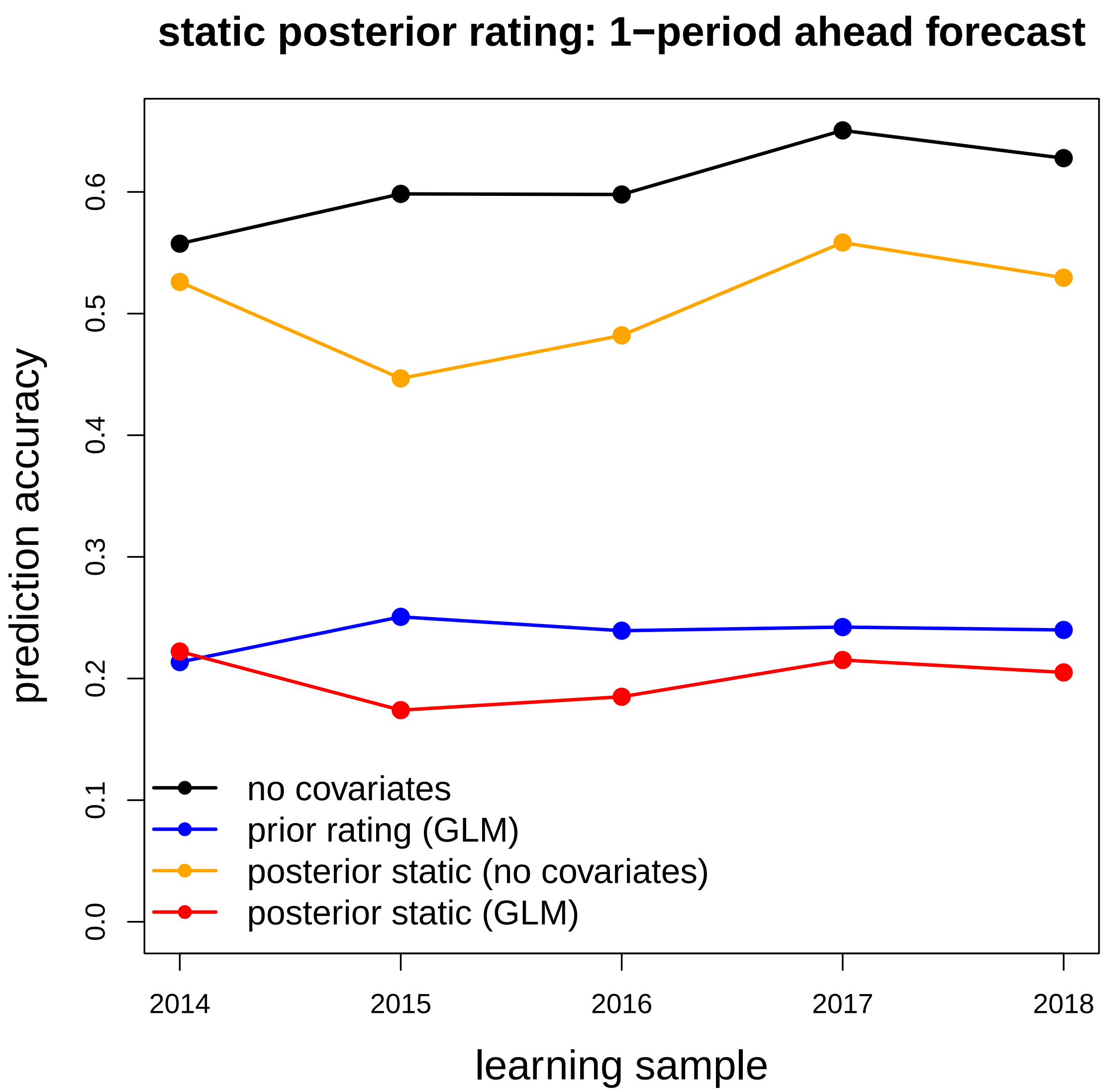
$$\omega_t = \frac{t}{t+\kappa} \qquad \text{and} \qquad \bar{Y}_t = \frac{1}{t}\sum_{s=1}^{t} Y_s,$$

and prior mean $\mu_0 = \mathbb{E}[Y_{t+1}]$ and credibility coefficient $\kappa \geq 0$.

- No seniority weighting of past claims Y_s; Pinquet et al. (2001).

- **Issue:** Static latent factor Θ makes past claims $Y_{1:t}$ exchangeable.

Accuracy of successive 1-period ahead forecasting



- **Section 3: Posterior rating: dynamic case**

Static vs. dynamic random effects

- Static random effects: Responses $(Y_t)_{t \geq 1}$ depend on static latent factor Θ.
- Dynamic random effects: Responses $(Y_t)_{t \geq 1}$ depend on latent process $(\Theta_t)_{t \geq 1}$.
- Best known dynamic random effects model: Kalman filter (1960) type

This model is *parameter-driven*, meaning that the model parameters fully specify the dynamics of the latent state-space process $(\Theta_t)_{t \geq 1}$.

Static vs. dynamic random effects

- Static random effects: Responses $(Y_t)_{t \geq 1}$ depend on static latent factor Θ.
- Dynamic random effects: Responses $(Y_t)_{t \geq 1}$ depend on latent process $(\Theta_t)_{t \geq 1}$.
- Best known dynamic random effects model: Kalman filter (1960) type

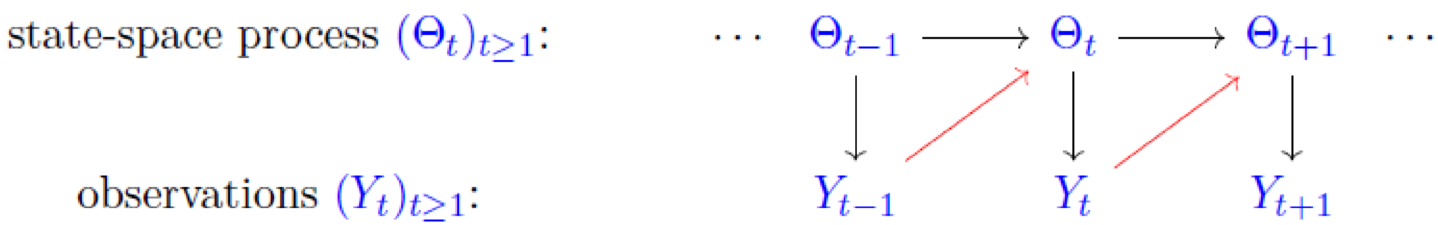
$$
\begin{array}{lccccccc}
\text{state-space process } (\Theta_t)_{t\geq 1}: & \cdots & \Theta_{t-1} & \longrightarrow & \Theta_t & \longrightarrow & \Theta_{t+1} & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & \\
\text{observations } (Y_t)_{t\geq 1}: & & Y_{t-1} & & Y_t & & Y_{t+1} &
\end{array}
$$

This model is *parameter-driven*, meaning that the model parameters fully specify the dynamics of the latent state-space process $(\Theta_t)_{t \geq 1}$.

Observation-driven dynamic random effects

- *Observation-driven* dynamic random effects models have been introduced by Harrison–Stevens (1976), Smith–Miller (1986), Harvey–Fernandes (1989).

- Observation-driven dynamic random effects models have a feedback loop:



- Harvey–Fernandes' (1989) proposal has an explosive long-term variance behavior.

- Ahn et al. (2023) extend this to different long-term variance behaviors in the Poisson-gamma conjugate prior case (this model is analytically tractable).

Poisson-gamma dynamic case (1/2)

(1) Observation equation:

$$Y_t|_{\{\Theta_{1:t}, Y_{1:t-1}\}} \sim \text{Poi}\,(\mu\Theta_t).$$

(2) Bayesian inference:

$$\Theta_t|_{\{\Theta_{1:t-1}, Y_t\}} \sim \Gamma\,(\alpha_t + Y_t, \beta_t + \mu).$$

(3) Transition equation (Kalman filter):

$$\Theta_{t+1}|_{\{\Theta_{1:t}, Y_{1:t}\}} \sim \Gamma\Big(\alpha_{t+1}(\Theta_{1:t}), \beta_{t+1}(\Theta_{1:t})\Big),$$

with scale and shape parameters β_{t+1} and α_{t+1}.

Poisson-gamma dynamic case (2/2)

(1) Observation equation:

$$Y_t|_{\{\Theta_{1:t}, Y_{1:t-1}\}} \sim \mathrm{Poi}\,(\mu\Theta_t).$$

(2) Bayesian inference:

$$\Theta_t|_{Y_{1:t}} \sim \Gamma\,(\alpha_t + Y_t, \beta_t + \mu).$$

(3) Observation-driven state-space update:

$$\Theta_{t+1}|_{Y_{1:t}} \sim \Gamma\Big(\alpha_{t+1}(Y_{1:t}),\ \beta_{t+1}\Big),$$

with deterministic scale β_{t+1} and shape parameter $\alpha_{t+1}(Y_{1:t})$.

Construction of step (3): state-space update

- Lukacs (1955): For independent random variables (with appropriate parameters)

$$\Theta \sim \Gamma \quad \text{and} \quad B \sim \mathrm{Beta} \quad \Longrightarrow \quad \Theta B \sim \Gamma.$$

  This allows for thinning in a gamma process.

- Observation-driven state-space update $\Theta_t \to \Theta_{t+1}$: Additionally, choose an independent gamma noise $\eta \sim \Gamma$ (with appropriate parameters)

$$\Theta_{t+1}|_{Y_{1:t}} = \frac{\Theta_t B}{q} + \eta \Big|_{Y_{1:t}} \sim \Gamma\left(\alpha_{t+1}, \beta_{t+1}\right),$$

  with parameter updates

$$\begin{aligned} \beta_t &\to \beta_{t+1} = q\left(\beta_t + \mu\right) > 0, \\ \alpha_t &\to \alpha_{t+1} = pq\left(\alpha_t + Y_t\right) + (1-p)\,\beta_{t+1} > 0. \end{aligned}$$

  for given constants $p, q \in (0, 1]$.

Construction of step (3): state-space update

- Lukacs (1955): For independent random variables (with appropriate parameters)

$$\Theta \sim \Gamma \quad \text{and} \quad B \sim \mathrm{Beta} \quad \Longrightarrow \quad \Theta B \sim \Gamma.$$

  This allows for thinning in a gamma process.

- Observation-driven state-space update $\Theta_t \to \Theta_{t+1}$: Additionally, choose an independent gamma noise $\eta \sim \Gamma$ (with appropriate parameters)

$$\Theta_{t+1}|_{Y_{1:t}} = \frac{\Theta_t B}{q} + \eta \Big|_{Y_{1:t}} \sim \Gamma\left(\alpha_{t+1}, \beta_{t+1}\right),$$

  with parameter updates

$$\begin{aligned} \beta_t &\to \beta_{t+1} = q\left(\beta_t + \mu\right) > 0, \\ \alpha_t &\to \alpha_{t+1} = pq\left(\alpha_t + Y_t\right) + (1-p)\,\beta_{t+1} > 0, \end{aligned}$$

  for given constants $p, q \in (0, 1]$.

Long-term behavior

- This model is mean-stationary: $\mathbb{E}[\Theta_t] = 1$ for all $t \geq 1$.

- **Explosive variance case:** $p = 1$ and $q < 1$

$$\lim_{t \to \infty} \mathrm{Var}(\Theta_t) = \infty.$$

- **Vanishing variance case:** $p < 1$ and $q = 1$

$$\lim_{t \to \infty} \mathrm{Var}(\Theta_t) = 0.$$

- **Bounded variance case:** $p < 1$ and $q < 1$

$$\inf_t \mathrm{Var}(\Theta_t) > 0 \qquad \text{and} \qquad \sup_t \mathrm{Var}(\Theta_t) < \infty.$$

Log-likelihood and model fitting

- The log-likelihood is fully tractable

$$\ell_{Y_{1:t}} = \sum_{s=1}^{t} \log \left(\frac{\Gamma\left(\alpha_s + Y_s\right)}{\Gamma\left(\alpha_s\right) Y_s!} \left(1 - \frac{\mu}{\beta_s + \mu}\right)^{\alpha_s} \left(\frac{\mu}{\beta_s + \mu}\right)^{Y_s} \right).$$

- These are negative binomial (marginal) models with $\alpha_s = \alpha_s(Y_{1:s-1})$.

- This is an integer-valued auto-regressive (INAR) negative binomial model.

- We can perform empirical Bayes' fitting.

Recursive credibility formula

- We get a closed form recursive experience rating formula

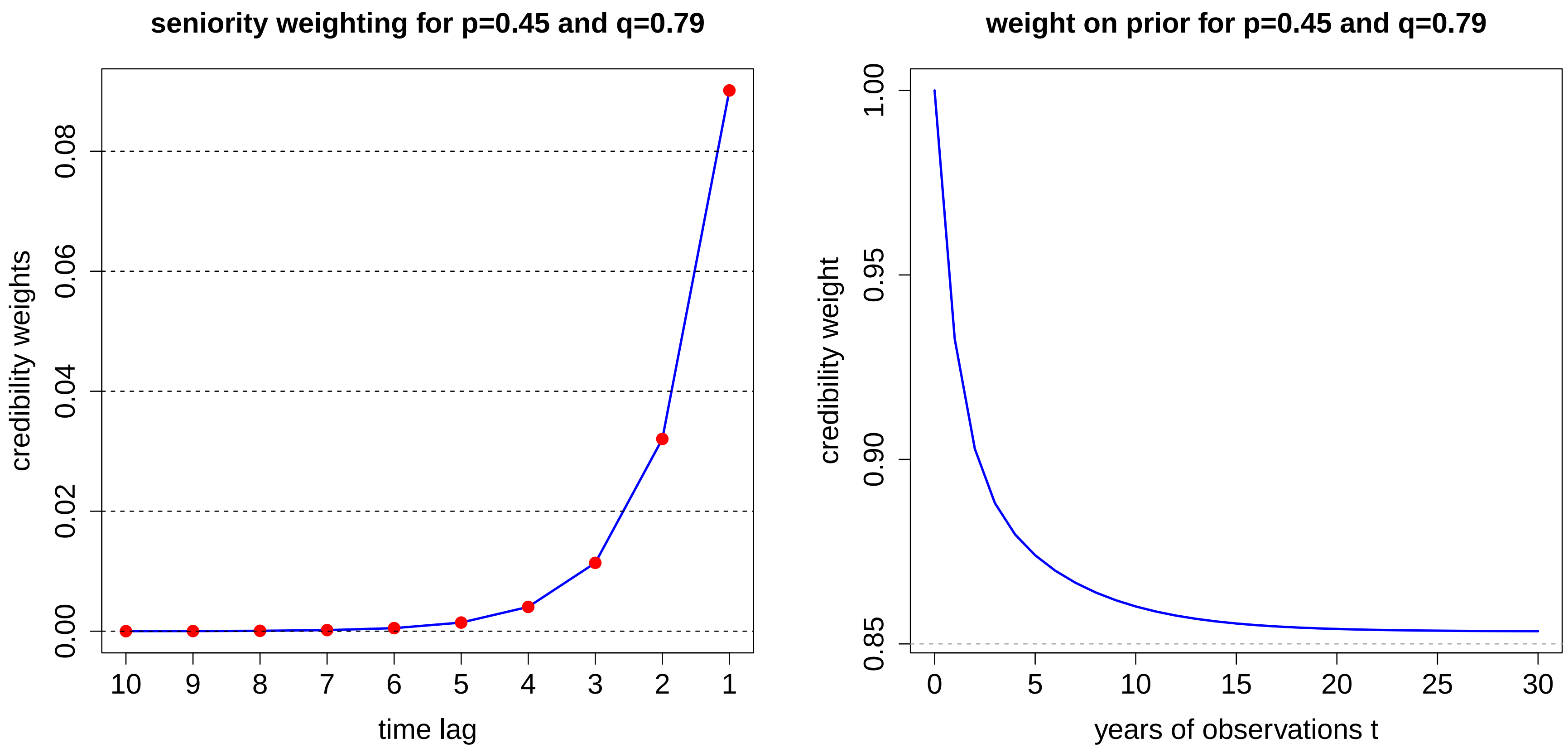
$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \mathbb{E}\left[Y_{t+1}\middle| Y_{1:t}\right] = \frac{\alpha_{t+1}}{\beta_{t+1}}\,\mu$$
$$= p\left(\omega_t\,Y_t + (1-\omega_t)\,\frac{\alpha_t}{\beta_t}\,\mu\right) + (1-p)\,\mu,$$

with (deterministic) credibility weights

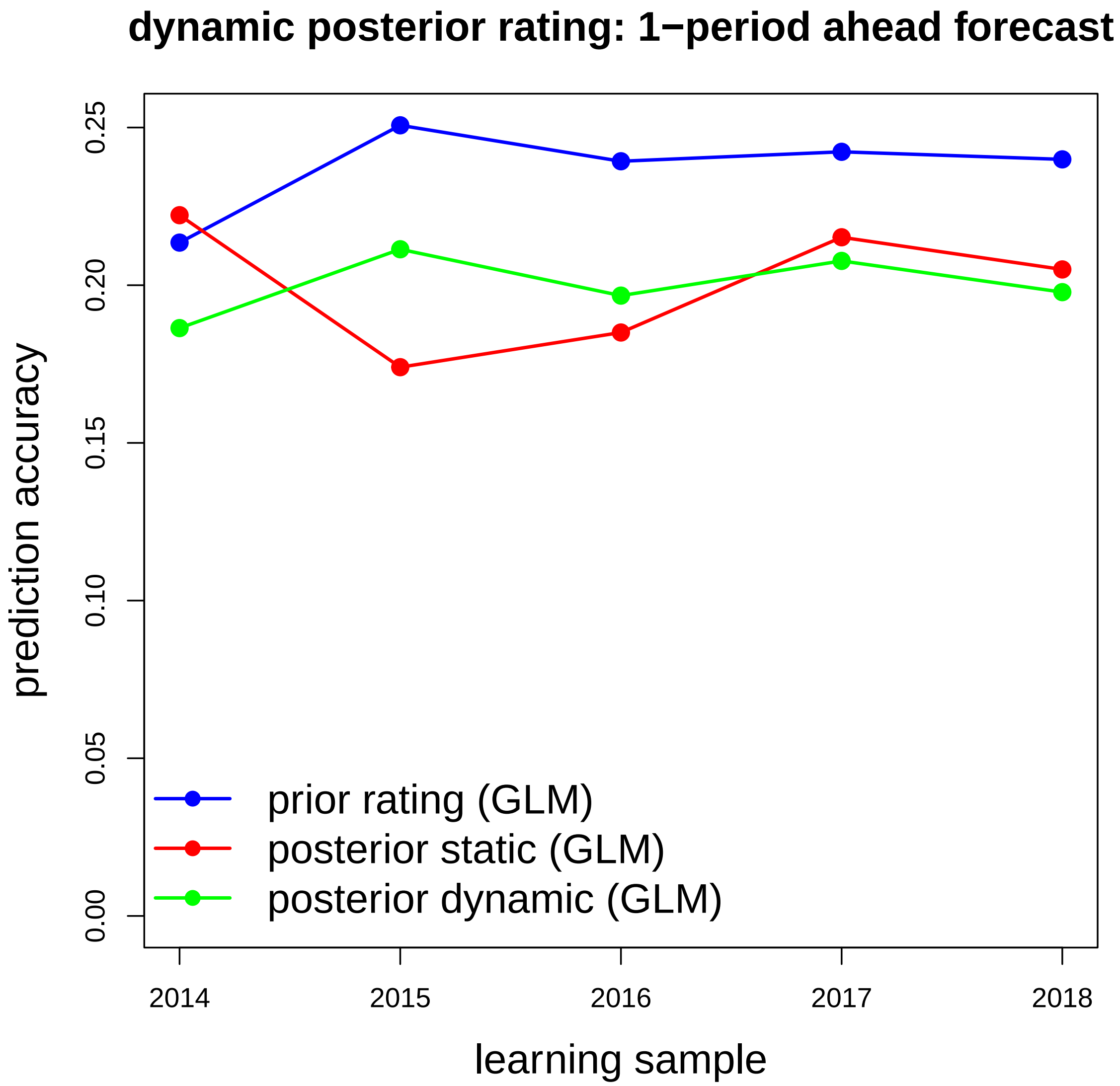
$$\omega_t = \frac{\mu}{\mu+\beta_t} \;\in\; (0,1).$$

- This provides seniority weighting of past claims, e.g., for $p < 1$.

Seniority weighting for $\widehat{p} = 0.45$ and $\widehat{q} = 0.79$



Accuracy of successive 1-period ahead forecasting



- **Section 4: Deep experience rating**

(Deep) attention weights

- Consider a linear (deep) attention approach

$$\mu^{\text{post}}_{Y_{t+1}|Y_{1:t}} = \sum_{s=1}^{t} \omega_{t,s} Y_s + \left(1 - \sum_{s=1}^{t} \omega_{t,s}\right) \mu(\boldsymbol{x}_{t+1}),$$

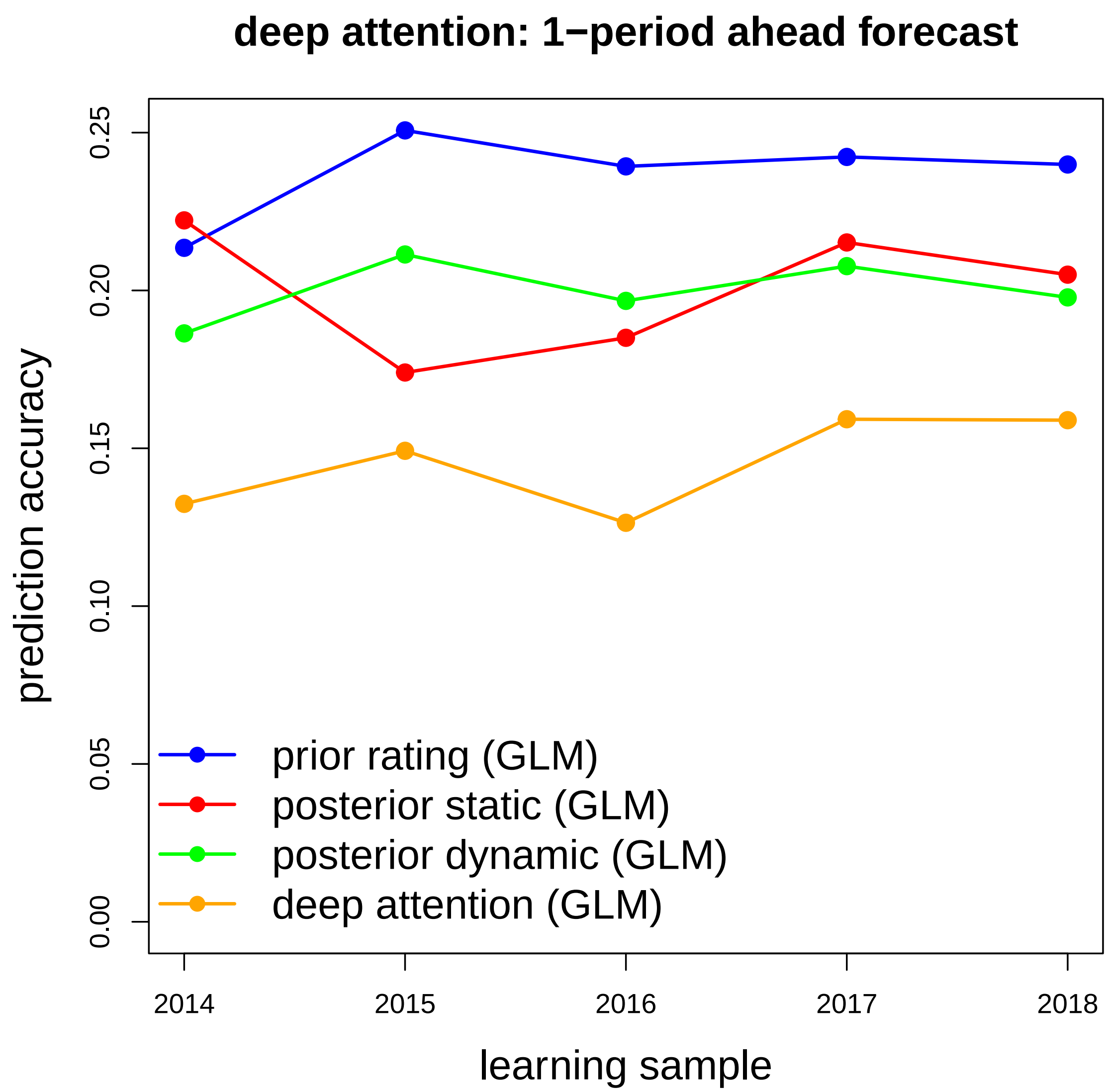
with (1-bounded) attention weights

$$\boldsymbol{x}_{1:t+1} \mapsto \omega_{t,s} = \omega_{t,s}(\boldsymbol{x}_{1:t+1}) \in (0,1).$$

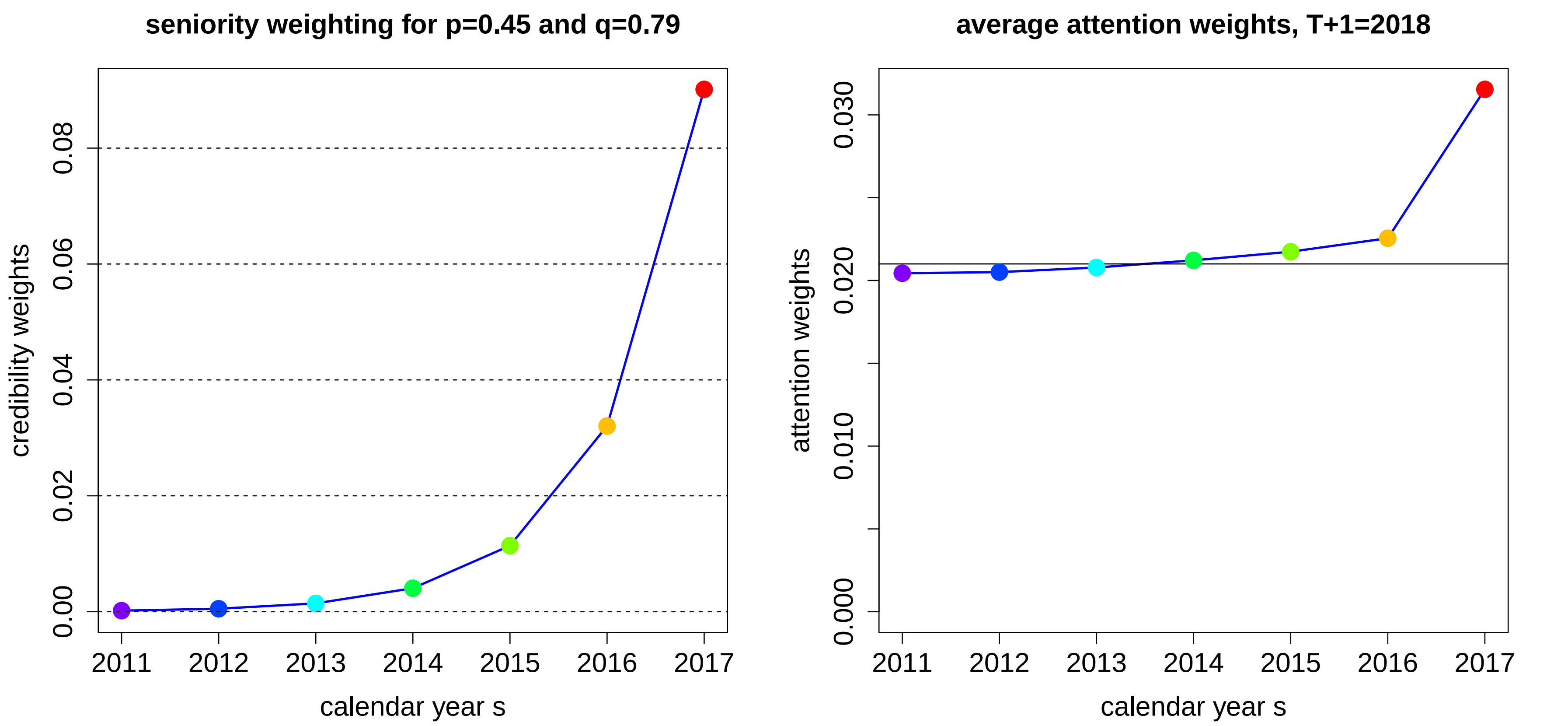
- This has the structure of an attention layer using a *key*, *query* and *value*; see Vaswani et al. (2017).

- This approach is distribution-free: fitting requires a strictly consistent loss function for mean estimation, see Gneiting (2011).

Accuracy of successive 1-period ahead forecasting



Seniority weighting of past claims



Conclusions

- Past claims have predictive power.
- Experience rating: past claims should receive a seniority weighting.
- Seniority weighting can be received in dynamic random effects models.
- There are tractable observation-driven dynamic random effects models.
- Distribution-free deep experience rating is based attention mechanisms.
- Attention mechanism also allows for non-linear credibility considerations.
- Distribution-free approaches require careful selection of objective functions for model fitting and mean estimation.
- We have only focused on predictive power and not on commercial pricing.

References

[1] Ahn, J.Y., Jeong H., Lu, Y., Wüthrich, M.V. (2023). A classification of observation-driven state-space count models for panel data. *arXiv*:2308.16058.

[2] Bichsel, F. (1964). Erfahrungstarifierung in der Motorfahrzeug-Haftpflicht-Versicherung. *Bulletin of the Swiss Association of Actuaries* **64**, 119-130.

[3] Bühlmann, H., Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries* **70**, 111-131.

[4] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106/494**, 746-762.

[5] Harrison, P.J., Stevens, C.F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society: Series B* **38/3**, 205-228.

[6] Harvey, A.C., Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics* **7/4**, 407-417.

[7] Jewell, W.S. (1974). Credible means are exact Bayesian for exponential families. *ASTIN Bulletin* **8**, 77-90.

[8] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82/1**, 35-45.

[9] Lukacs, E. (1955). A characterization of gamma distribution. *The Annals of Mathematical Statistics* **26/2**, 319-324.

[10] Pinquet, J., Guillén, M., Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin* **31/2**, 337-348.

[11] Smith, R.L., Miller, J.E. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society. Series B (Methodological)*. **48/1**, 79-88.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. *arXiv*:1706.03762v5

[13] Wüthrich, M.V. (2024). *Experience Rating for Insurance Pricing*. SSRN Manuscript ID 4726206.