Business School

UNSW AUSTRALIA

Never Stand Still | Business School | Risk and Actuarial Studies

The long road to enlightenment

Loss reserving models from the past, with some speculation on the future

Greg Taylor
School of Risk and Actuarial Studies
University of New South Wales
Sydney, Australia

Advances in predictive analytics - with applications in insurance and risk management
University of Waterloo, Waterloo, Ontario, Canada December 1-2, 2017

UNSW AUSTRALIA | AGSM
Business School

Overview

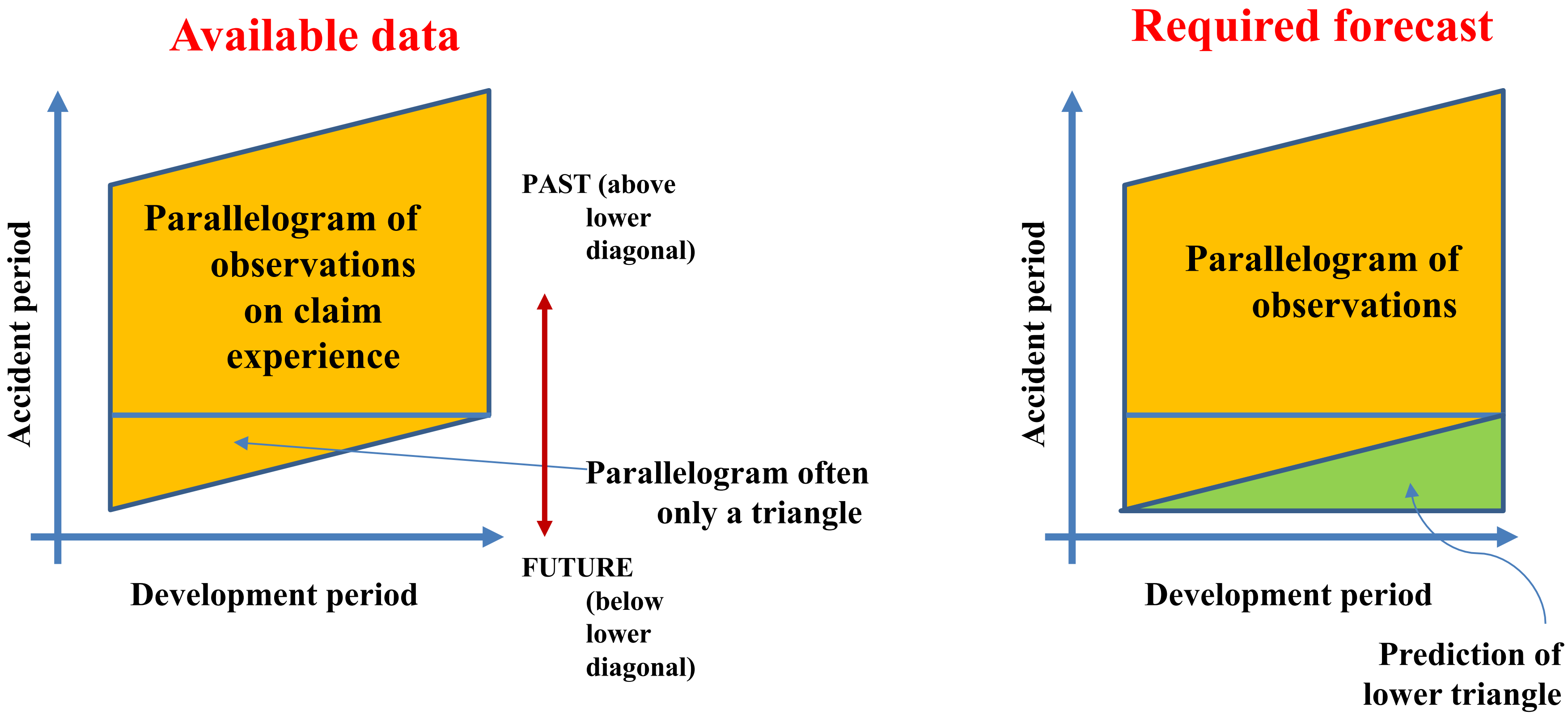
- The Landscape
- The Jurassic period
  – Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  – Stochastic models
- The Paleogene – increased diversity in the higher forms
  – Evolutionary models
  – Parameter reduction
  – Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  – Artificial intelligence
- The later Anthropocene – the future

Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

The Landscape

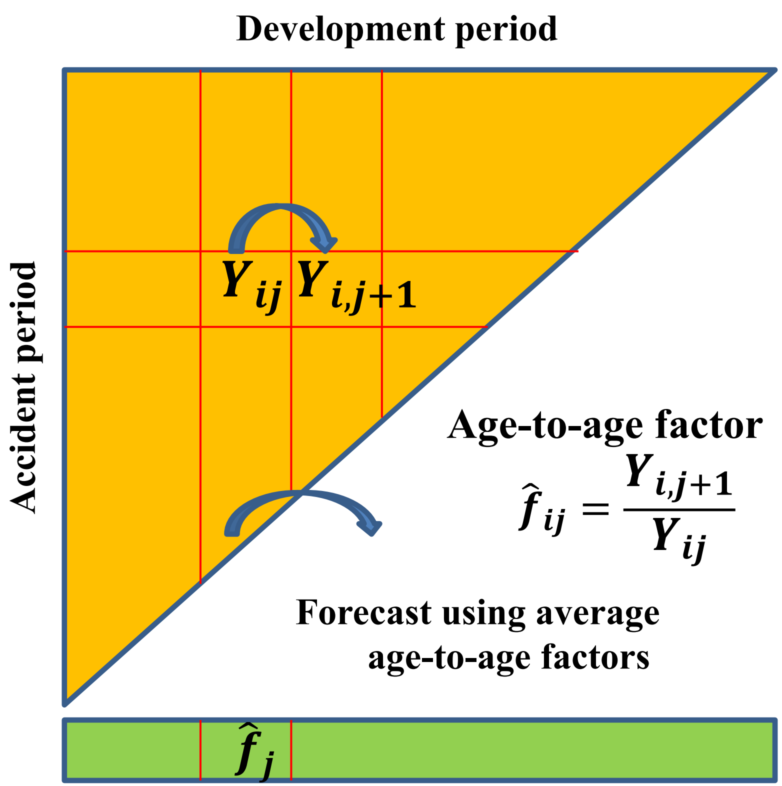
- Loss reserving
  - A problem in forecasting



Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

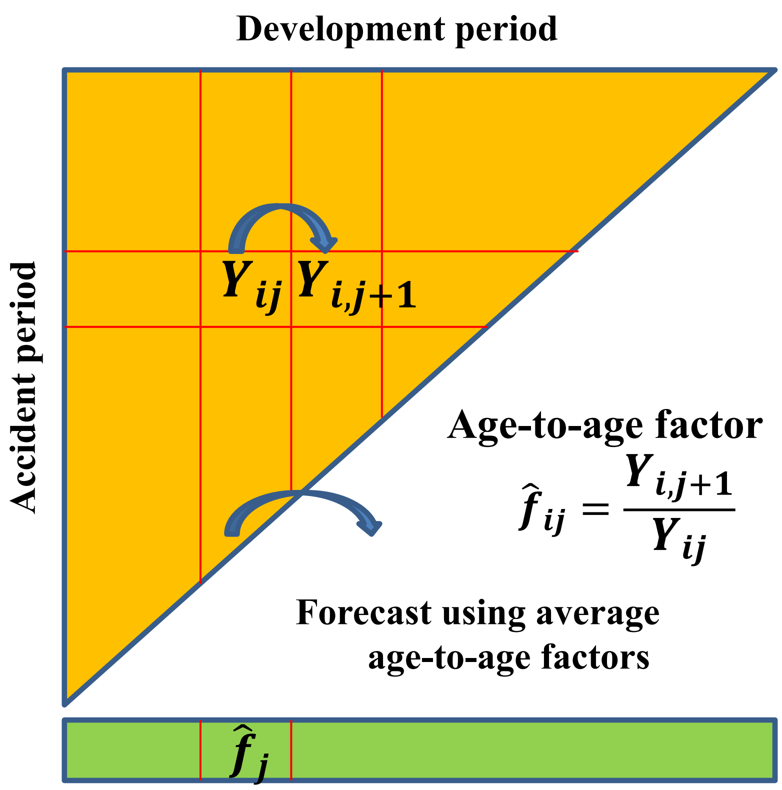
Earliest “models”



Some form of averaging of age-to age factors over each row

- Observations organized by row and column
- An accident period (row) “develops” from one column to the next
- Forecast is $Y_{i,j+1} = \hat{f}_j Y_{ij}$
- These “models” include (Taylor, 1986, 2000; Wüthrich & Merz, 2008):
  - Chain ladder
  - Separation method
  - and all their derivatives:
    - Bornhuetter-Ferguson
    - Cape Cod;
    - etc.

Properties of earliest "models"

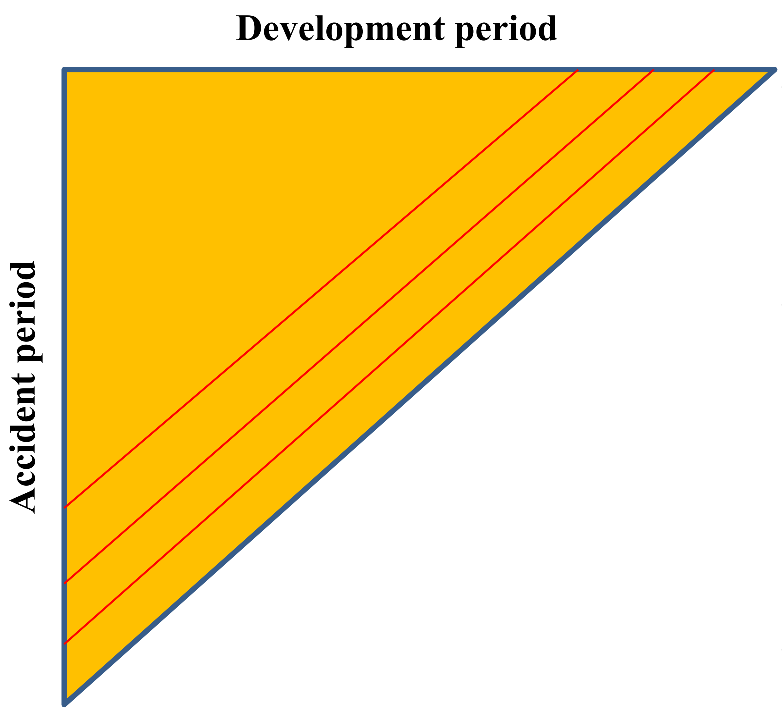


Some form of averaging of age-to age factors over each row

- No model formulated
  - No stochastics introduced
  - Actually a "procedure" or "algorithm" rather than model
- Assumption of same age-to-age factors for each row
- Parameter estimation carried out by row, column or diagonal averaging
- In statistical parlance, include "row, column and/or diagonal effects"
- Over-parameterized
  - For an $n \times n$ triangle, chain ladder involves $2n - 1$ parameters
  - This increases prediction error

UNSW AUSTRALIA

Unfitness of Jurassic denizens



- Most early models include row and column effects
- What if there is a need to include diagonal effects also?
  - e.g. variable inflation
- What if rate of claim settlement changes from one row to another (Fisher & Lange, 1973)?
  - Age-to-age factors vary from row to row
- Such features:
  - Increase parameterization
  - Are difficult to parameterize by row/column/ diagonal manipulation (Taylor, 2000)

Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

The appearance of stochastic models

- The Jurassic models were unformulated, but essentially took the form:

Observation set
Parameter set

$$Y_{ij} = f(Y, \alpha)$$

- This is not stochastic, but can easily be made so:

$$Y_{ij} = f(Y, \alpha) + \varepsilon_{ij}, \qquad E[\varepsilon_{ij}] = 0$$

- With some restriction on f and ε_{ij}, one arrives at a GLM

$$Y_{ij} \sim F(\mu_{ij}, \varphi / w_{ij}), \qquad \mu_{ij} = E[Y_{ij}], \qquad h(\mu_{ij}) = x_{ij}^T \beta,$$

$$F \in EDF$$

Exponential dispersion family

Scale parameter

Link function

Covariate vector

Parameter vector

- Chain ladder example ($Y_{i,j+1} = \hat{f}_j Y_{ij}$)

$$F = ODP, h = \ln, \quad x_{i,j+1} = [0, \ldots 0, Y_{ij}, 0 \ldots 0], \beta = [f_1, f_2, \ldots]^T$$

Brief history of stochastic models

- Notably advanced creatures of the Jurassic were:
  - Stochastic claims analysis (Reid, 1978)
  - A stochastic chain ladder model (Hachemeister & Stanard, 1975)
  - An individual claim development model (Hachemeister, 1978, 1980)
- History of actuarial GLMs longer than often realized:
  - 1972: concept introduced (Nelder & Wedderburn)
  - 1977: GLIM software introduced
  - 1984: Tweedie family introduced (Tweedie, 1984)
  - 1990+: seminal actuarial papers (Wright, 1990; Brockman & Wright, 1992)
  - Note, however:
    - Early application of GLMs to pricing (Baxter, Coutts & Ross, 1979)
    - Use within my own consulting practices through the 1980's

More recent loss reserving GLMs

- Used to model claim data sets with many complex and overlapping features, e.g.
  - Taylor & McGuire (2004)
    - Auto liability
    - Rates of claim settlement vary over time
    - Superimposed inflation varies with payment quarter and operational time
    - Legislative change (accident quarter)
  - Taylor & Mulquiney (2007)
    - Mortgage insurance
    - Cascaded model with sub-models for healthy policies, in arrears, properties in possession, and claims
  - Taylor, McGuire & Sullivan (2008)
    - Medical malpractice
    - Individual claim development model with covariates such as specialty, geographic area of practice, etc.
  - Taylor & McGuire (2016) – a monograph on GLM reserving
- This type of analysis is now called **Predictive Analytics**

Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

Adaptation of species

- GLM model was

$$Y_{ij} \sim F\left(\mu_{ij}, \varphi/w_{ij}\right), \qquad \mu_{ij} = E\left[Y_{ij}\right], \qquad h\left(\mu_{ij}\right) = x_{ij}^T \beta$$

- Here parameter set β is constant over time
  - What if it is expected to change?
- One can introduce an **evolutionary model** in which parameters vary over time, e.g. (with **time** $t = i + j$)

$$Y_{ij} \sim F\left(\mu_{ij}^{(t)}, \varphi/w_{ij}\right), \qquad \mu_{ij}^{(t)} = E\left[Y_{ij}\right], \qquad h\left(\mu_{ij}^{(t)}\right) = x_{ij}^T \beta^{(t)}$$

$$\beta^{(t)} \sim P\left(.; \beta^{(t-1)}, \psi\right)$$

Conjugate prior

Prior dispersion structure

- See
  - Taylor (2008)
  - Taylor & McGuire (2009)

Adaptive reserving (1)

- Adaptive model

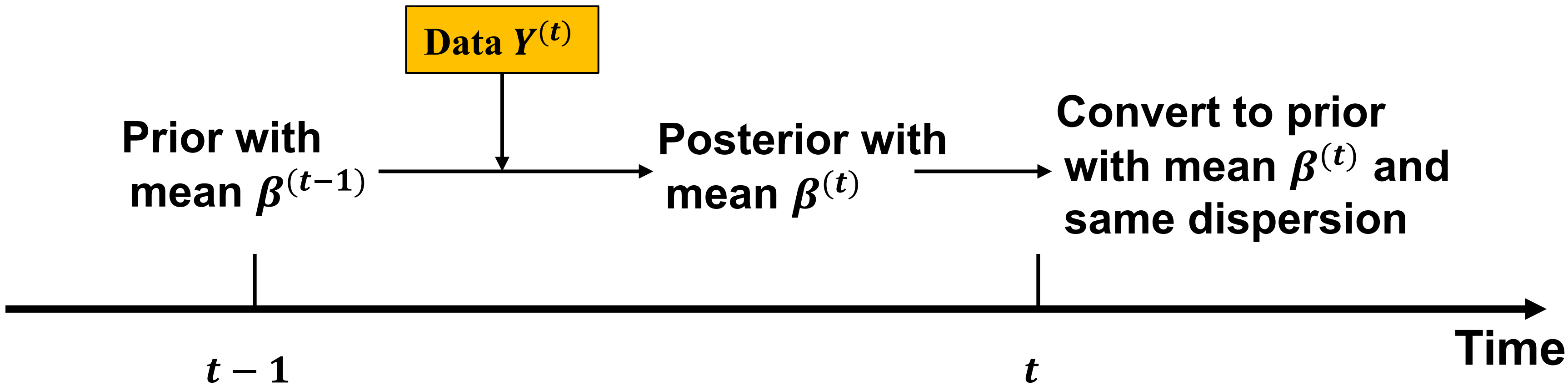
$$Y_{ij} \sim F\left(\mu_{ij}^{(t)}, \varphi/w_{ij}\right), \qquad \mu_{ij}^{(t)} = E[Y_{ij}], \qquad h\left(\mu_{ij}^{(t)}\right) = x_{ij}^T \beta^{(t)}$$

$$\beta^{(t)} \sim P\left(.; \beta^{(t-1)}, \psi\right)$$

- Reminiscent of **Kalman filter** (see Harvey, 1989), **BUT**
  - KF applies to linear models, whereas this one is **non-linear** in general
  - KF assumes normal error for observations, whereas this model assumes **non-normal**
  - The posterior likelihood at each t does not lie within the set of EDF conjugate priors
    - Must be approximated by a member of the set with same first and second order moments
    - Some stability problems

Adaptive reserving (2)

- Schematic of process from time $t-1$ to t



UNSW AUSTRALIA

Miniaturization: dimensionality reduction

- The Jurassic models were lumbering, with overblown parameter sets
- GLMs were more efficient but without much systematic attention to the issue
- A more recent approach that brings the issue into focus is **regularized regression**, and specifically the **least absolute shrinkage and selection operator (LASSO) model** (Tibshirani, 1996)

UNSW AUSTRALIA

Regularized regression

Replace squared error by GLM loss function (log-likelihood) to obtain regularized GLM

- Consider first linear regression, as opposed to GLM, and consider the loss function (in an obvious notation)

$$L(y;\beta) = \|y - X\beta\|_2^2 + \lambda\|\beta\|_p$$

where $\|.\|_p$ denotes the L_p norm and $\lambda > 0$ is a constant

This is **regularized (linear) regression**

- Note that
  - $\lambda = 0$ yields OLS regression
  - $\lambda \neq 0, p = 2$ yields Ridge regression
  - $\lambda \neq 0, p = 1$ yields the lasso

$\lambda \to 0$: no elimination of covariates

$\lambda \to \infty$: maximum elimination of covariates

- A property of the lasso is that it can force many components of β to zero
  - Thus an effective tool for elimination of covariates from a large set

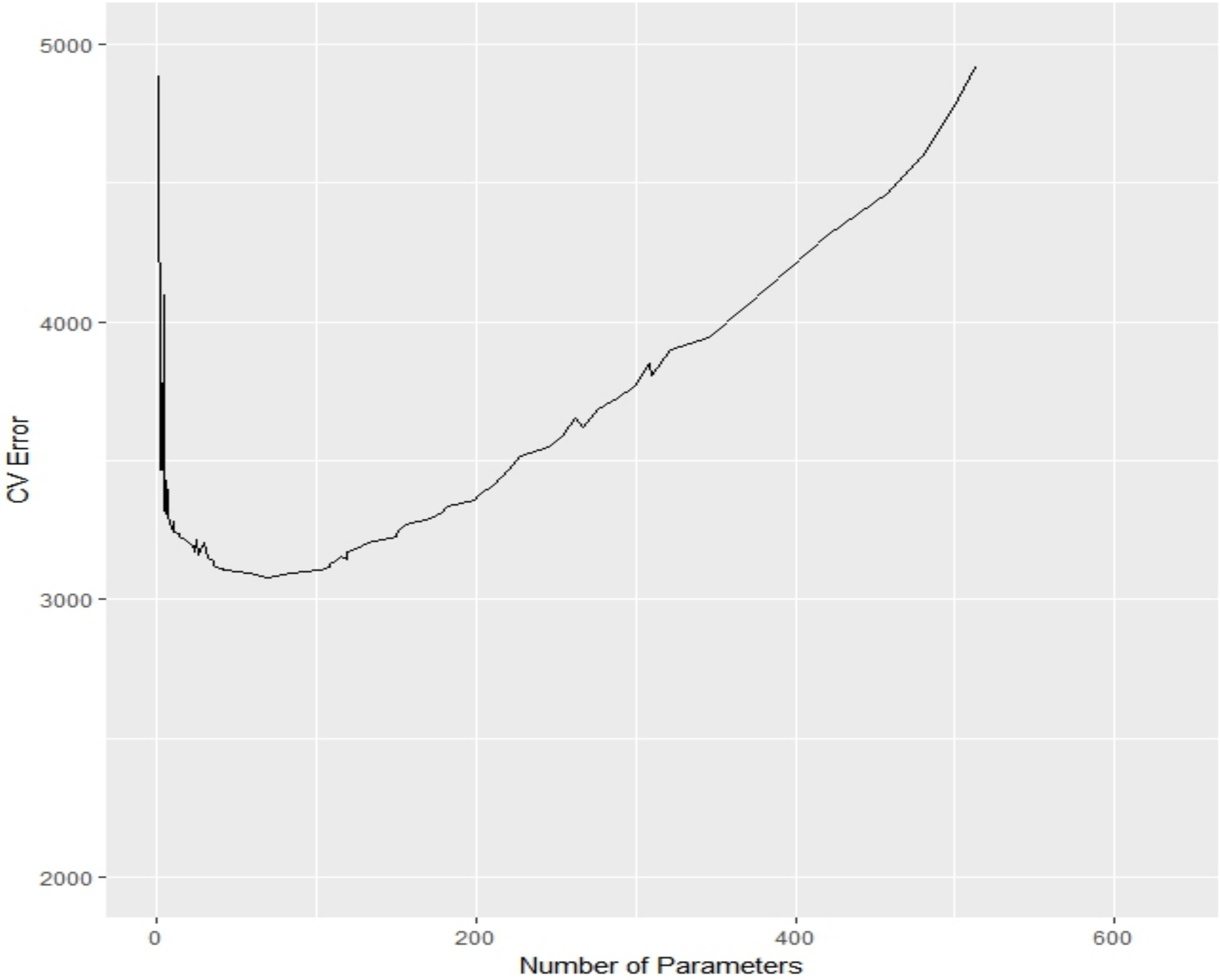
Calibration

- Consider a large set of basis functions from which all functions in the loss reserving model may be expressed as linear combinations
- Lasso used to reduce large covariate set to just the "essential" members
  - Sequence of models examined with increasing λ
  - Number of covariates decreasing
  - Model chosen to minimize cross-validation error
  - Examples
    - Venter & Şahîn (2017) – mortality
    - Gao & Meng(2017) – loss reserving
    - Taylor, McGuire & Miller (2016) – loss reserving

Cross-validation

- Randomly delete one n-th of the data set, as a test sample
- Fit the model to the remainder of the data set (training set)
- Generate fitted values for the test sample
- Compute error between test sample and fitted values (e.g. sum of squares)

Example of lasso calibration



Forecast error (1)

- Let
  - R denote the amount of unpaid losses (a random variable)
  - $\hat{R}$ denote an estimate of R (assumed unbiased)
- One wishes to know something of the distribution of $\hat{R}$, e.g.
  - The full distribution
  - Certain quantiles (risk margins, capital margins)
  - Just the mean square error of prediction (MSEP): $E\left[\left(R-\hat{R}\right)^2\right]$
- If one is not concerned with the tails of the distribution (e.g. 75-percentile risk margin), then MSEP will often provide a measure of the forecast quality

Forecast error (2)

- There are two main approaches to the estimation of forecast error
  - Bootstrap
  - Markov Chain Monte Carlo (MCMC) (Meyers, 2015)
    - Relevant to Bayesian models
- Both estimate full distribution, and therefore any property of the distribution

Parameterization and forecast error

- Beyond a certain threshold, the inclusion of additional parameters in a model will result in over-fitting and increase MSEP
- Similar considerations apply to cascaded models (i.e. those involving multiple sub-models)
- Taylor & Xu (2016) investigated, for certain data sets,
  - Chain ladder (a single model involving only paid amounts); and
  - An alternative model, incorporating reported and finalized claim count information, and comprising 3 sub-models
  - The results indicated that the alternative produced lower MSEP when the data set failed to conform with the chain ladder parametric structure by a sufficient margin

The fine detail: granular (micro-) reserving

- Models the detail of individual claims, e.g.
  - Reporting date
  - Individual payment dates
  - Amounts of individual payments
- Generally regarded as commencing with Norberg (1993, 1999), Hesselager (1994), with implementation by Pigeon, Antonio & Denuit (2013, 2014) and Antonio & Plat (2014)
  - Note, however, the earlier implementations (Hachemeister (1978,1980), Taylor & Campbell (2002))
- Distinction between aggregate and granular models is largely false
  - Any model that includes claim counts can be regarded as granular
    - It produces forecasts at an individual claim level
    - Only a question of the volume of conditioning data
  - So one should think in terms of a aggregate-granular spectrum

Applications of granular reserving

- Loss reserving at the individual claim level has an application when loss reserves are required in respect of small groups of claims and physical estimates do not exist
  - e.g. in relation to small cost centres of an organization
- Otherwise, required only if they produce a loss reserve superior to that produced by aggregate methods
  - Recall that (aggregate) chain ladder is minimum variance for ODP observations (Taylor, 2011)
  - And remember that granular models will usually be cascaded
    - With their property of inflating prediction error
- Huang, Wu & Zhou (2016) claim that micro-models outperform aggregate
  - But their calibration and forecast are essentially the same as the **Payment per Claim Finalized** aggregate model found in the literature (Taylor, 1986, 2000)
  - Just conditioned by more data than their aggregate models
- So jury still out on the value of micro-models!

UNSW AUSTRALIA

Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

The rise of roboticus sapiens (1)

- First steps in machine learning
  - **Artificial neural nets (ANNs)**
    - Mulquiney (2006)
      - Modelled a set of claim finalizations tabulated by:
        - » Accident quarter
        - » Development quarter
        - » Payment quarter
        - » Operational time at finalization
        - » Season of finalization (calendar quarter)
      - ANN goodness-of-fit superior to GLM
      - ANN detected superimposed inflation that varied over both finalization quarter and operational time
      - ANN detected effects of a legislative change (accident quarter effect) that occurred in the midst of the claim experience

The rise of roboticus sapiens (2)

- Harej, Gächter & Jamal (2017)
  - IAA Working Group on "Individual Claim Development with Machine Learning"
  - This was an "under-powered" ANN which assumed chain ladder models for paid and incurred costs respectively for individual claims, and simply estimated the age-to-age factors
  - However, since it included both paid and incurred amounts, it managed to differentiate age-to-age factors for different claims
    - » e.g. claims with small amount paid but large amounts incurred showed high development of payments
- Wüthrich & Buser (2017) have produced a set of lecture notes on machine learning:
  - Regression trees
  - Random forests
  - Support vector machines
  - Clustering for telematics data

The watchmaker and the oracle (1)

- The tendency of micro-modelling (watchmaking) is to increase the number of cascaded sub-models
  - → individual claims
  - → individual payments, etc.
- Many parameters, with implications for prediction error
- Increases the fragility of the model
  - Increased complexity due to dependencies, e.g.
    - In Liability business, occurrence of a large payment would reduce the likelihood of another large payment
    - In Workers Compensation, a return to work from incapacity would usually lower the likelihood of immediate incapacity onset
    - Dependencies between sub-models render validation difficult
      - One may validate all sub-models internally, but then discover that the total model does not validate
- On the other hand, all aspects of the model are understood

The watchmaker and the oracle (2)

- ANN (oracle) is a model that observes all the complexity of the training data, and should accommodate it
  - By-passes all the difficulties of micro-modelling
- However, it is an extremely opaque model
  - At its core (the neurons), it consists of just a set of weighted averages
  - Individual data features (e.g. superimposed inflation) are hidden within the model
  - They may also be poorly measured
    - e.g. diagonal effects may be inaccurately measured , but compensated by measured, but actually non-existent, row effects
  - Can be difficult to validate
    - What is one's recourse in the event of validation failure?

Overview

- The Landscape
- The Jurassic period
  - Unformulated models
- The Cretaceous period – seed-bearing organisms appear
  - Stochastic models
- The Paleogene – increased diversity in the higher forms
  - Evolutionary models
  - Parameter reduction
  - Granular (micro-) models
- The Anthropocene period – intelligent beings intervene
  - Artificial intelligence
- The later Anthropocene – the future

The future?

- Aggregate models?
- Micro-modelling?
- Machine learning?

References (1)

- Antonio K & Plat R (2014). Micro-level stochastic loss reserving for general insurance, **Scandinavian Actuarial Journal**, 2014(7), 649-669.
- Baxter L A, Coutts S M & Ross S A F (1979). Applications of linear models in motor insurance. **21st International Congress of Actuaries**, 2, 11.
- Brockman M J & Wright T S (1992). Statistical motor rating: making effective use of your data. **Journal of the Institute of Actuaries**, 119, 457-526.
- Fisher & Lange 1973.
- Gao G & Meng S (2017). Stochastic claims reserving via a Bayesian spline model with random loss ratio effects. **Astin Bulletin**, On-line, 1-34.

References (2)

- Hachemeister C A (1978). A structural model for the analysis of loss reserves. **Bulletin d'Association Royal des Actuaires Belges**, 73, 17-27.
- Hachemeister C A (1980). A stochastic model for loss reserving. **Transactions of the 21st International Cogress of Actuaries**, 1, 185-194.
- Hachemeister C A & Stanard J N (1975). IBNR claims count estimation with static lag functions. **Spring meeting of the Casualty Actuarial Society**.
- Harej B, Gächter R & Jamal S (2017). Report of the International Actuarial Association Working Group on "Individual Claim Development with Machine Learning". http://www.actuaries.org/ASTIN/Documents/ASTIN_ICDML_WP_Report_final.pdf.

UNSW AUSTRALIA

References (3)

- Harvey A C (1989). **Forecasting, structural time series and the Kalman filter**. Cambridge University Press, Cambridge, UK.
- Hesselager O (1994). A Markov Model for Loss Reserving. **Astin Bulletin**, 24(2), 183-193.
- Huang J, Wu X & Zhou X. (2016). Asymptotic behaviors of stochastic reserving: aggregate versus individual models. **European Journal of Operational Research**, 249, 657-666.
- Meyers G (2015). Stochastic loss reserving using Bayesian MCMC models. **CAS Monograph Series, number 1**. Monograph commissioned by the Casualty Actuarial Society, Arlington VA, USA.
- Mulquiney P (2006). Artificial neural networks in insurance loss reserving. 9th Joint Conference on Information Sciences 2006 – Proceedings. Atlantis Press. http://www.atlantis-press.com/php/download_paper?id=67.

References (3)

- Nelder J A & Wedderburn R W M (1972). Generalised linear models. **Journal of the Royal Statistical Society, Series A**, 135, 370-384.
- Norberg R (1993). Prediction of outstanding liabilities in non-life insurance. **Astin Bulletin**, 23(1), 95–115.
- Norberg R (1999). Prediction of outstanding liabilities II. Model extensions variations and extensions. **ASTIN Bulletin**, 29(1), 5–25.
- Reid 1980
- Pigeon M, Antonio K & Denuit M (2013). Individual loss reserving with the multivariate skew normal framework. **Astin Bulletin**, 43(3), 399-428.
- Pigeon M, Antonio K & Denuit M (2014). Individual loss reserving using paid–incurred data. **Insurance: mathematics and economics**, 58, 121-131.

UNSW AUSTRALIA

References (4)

- Taylor G (2000). **Loss reserving: an actuarial perspective**. Kluwer Academic Publishers, Dordrecht, Netherlands.
- Taylor G (2008). Second order Bayesian revision of a generalised linear model. **Scandinavian Actuarial Journal,** 2008(4), 202-242.
- Taylor G (2011). Maximum likelihood and estimation efficiency of the chain ladder. **Astin Bulletin**, 41(1), 131-155.
- Taylor G C (1986). **Claims reserving in non-life insurance**. North-Holland, Amsterdam.
- Taylor G & Campbell M (2002). Statistical case estimation. **Research Paper No 104 of the Centre for Actuarial Studies, University of Melbourne**. Appears at http://fbe.unimelb.edu.au/__data/assets/pdf_file/0005/806396/104.pdf

References (5)

- Taylor G & McGuire G (2004). Loss reserving with GLMs: a case study. (with McGuire, G) **Casualty Actuarial Society 2004 Discussion Paper Program**, 327-392.
- Taylor G & McGuire G (2009). Adaptive reserving using Bayesian revision for the exponential dispersion family. **Variance**, 3, 105-130.
- Taylor G & McGuire G (2016). Stochastic loss reserving using Generalized Linear Models. **CAS Monograph Series, number 3**. Monograph commissioned by the Casualty Actuarial Society, Arlington VA, USA.
- Taylor G, McGuire G & Sullivan J (2008). Individual claim loss reserving conditioned by case estimates (with G McGuire and J Sullivan). **Annals of Actuarial Science**, 3 (2008(1&2)), 215-256.
- Taylor G & Mulquiney P (2007). Modelling mortgage insurance as a multi-state process. **Variance**, 1, 81-102.

References (6)

- Taylor G & Xu J (2016). An empirical investigation of the value of finalisation count information to loss reserving. **Variance**, 10(1), 75-120.
- Tibshirani R (1996). Regression Shrinkage and Selection via the lasso. **Journal of the Royal Statistical Society, Series B**, 58 (1), 267–88.
- Tweedie M C K (1984). An index which distinguishes between some important exponential families, **Statistics: Applications and New Directions, Proceedings of the Indian Statistical Golden Jubilee International Conference**, J. K. Ghosh and J. Roy (Eds.), Indian Statistical Institute, 1984, 579-604.
- Venter G & Şahîn S (2017). Parsimonious parameterization of age-period-cohort models by Bayesian shrinkage. **Astin Bulletin**, 47(1), 1-22.

References (7)

- Wüthrich M V & Buser C (2017). **Data analytics for non-life insurance pricing**. RiskLab Switzerland, Department of Mathematics, ETH Zurich.
- Wüthrich M V & Merz M (2008). **Stochastic claim reserving methods in insurance**. John Wiley & Sons, Ltd, Chichester, UK.
- Wright T S (1990). A stochastic method for claims reserving in general insurance. **Journal of the Institute of Actuaries**, 117, 677-731.