Expanding Horizons

October 2018
Issue 57

The Long Road to Enlightenment:
Loss Reserving Models from the Past, with Some Speculation on the Future

By Greg Taylor

Editor's Note: This article is the first in a new series in Expanding Horizons *on "The Art of Actuarial Science."*

1. The Jurassic Period

Everyone is familiar with the earliest find in the fossil record, the **chain ladder**. Though perhaps of mysterious provenance, it has been the mainstay of loss reserving for many years and, to a large degree, reliance on it continues.

In its earliest form, it consisted of the computation of **age-to-age factors**, defined as

$$f_j = \bar{Y}_{j+1}/\bar{Y}_j \qquad (1.1)$$

where $\bar{Y}_j$ is some kind of average of observations Y_{ij} for accident period i and development period j, and $\bar{Y}_{j+1}$ is the comparable average for development period j+1. For example, the observation Y_{ij} might denote cumulative claim payments to the end of development period j.

It was supposed that the factor f_j measured the development from j to $j+1$. This measurement was taken over certain accident periods, and then assumed to apply to others for which development from j to $j+1$ lay in the future. In this way forecasts of future claim experience, and hence loss reserves were formulated.

The same evolutionary branch contained other variants of this creature, such as the **Bornhuetter-Ferguson** procedure (Bornhuetter & Ferguson, 1972), **Cape Cod** (Bühlmann, 1983, 2016), **Benktander-Hovinen** (Benktander, 1976; Hovinen, 1981), and so on.

In their original skeletal forms, such as described above for the chain ladder, none of these approaches could fairly be referred to as a "model." They were calculation routines, or algorithms, but this alone did not qualify them as models.

According to more modern convention, a model requires a set of statements describing the stochastic properties of the observations, considered as random variables. For example, these properties might include specification of a set of model parameters, and how the mean of each observation is related to these parameters. It might include a statement on any stochastic dependencies between observations.

Ultimately, each of these primitive creatures evolved into a more complex life form, consisting of a stochastic model from which well-known forecasts could be derived by rigorous procedure, e.g., maximum likelihood. However, there are other reasons to question the evolutionary fitness of the denizens of the Jurassic world, reasons which are related to the simplicity of these models.

The early models included a column effect, as illustrated in (1.1), and also a row effect, dependent on i. But what if it were considered that data were subject to a diagonal effect, characterized by $i + j$, e.g., superimposed inflation? The models provided no obvious way forward.

What if the rate of claim settlement changes from one row to another (Fisher & Lange, 1973), so that the age-to-age factors from past accident periods are not representative of the future? Such features increase parameterization in models that are already crowded with parameters. Their accommodation by means of simple row/column/diagonal manipulation is difficult in any event.

2. The Cretaceous Period: Seed-Bearing Organisms Appear

Jurassic models were gradually succeeded by properly formulated stochastic models. These generally take the form

$$Y_{ij} = f(Y, \alpha) + \varepsilon_{ij}, E[\varepsilon_{ij}] = 0 \tag{2.1}$$

where f is some real-valued function, Y denotes the entire set of observations $Y_{k\ell}$, α some set of unknown parameters, and ε_{ij} a stochastic error.

The chain ladder (1.1) can be extended to a stochastic model in this form. Hachemeister & Stanard (1975) introduced an alternative stochastic representation of the chain ladder. Reid (1978) introduced a stochastic model of the payments in respect of individual claims, and Hachemeister (1978) a stochastic model of development of individual claim incurred costs, including manual estimates.

Many models of this sort can be accommodated within the framework of a **Generalized Linear Model (GLM)**. These are models of form (2.1) but with the distribution of the Y_{ij} restricted to the **exponential dispersion family (EDF)** (Nelder & Wedderburn, 1972).

This family was convenient because the software package GLIM was introduced in 1977 for parameter estimation. The **Tweedie sub-family** of the EDF was introduced in Tweedie (1984), and a number of other GLM packages introduced for models from this more restricted (but still extensive) family. Seminal papers (Wright, 1990; Brockman & Wright, 1992) cemented GLMs in the actuarial literature. Note, however, the much earlier application (Baxter, Coutts & Ross, 1979).

The GLM framework has proven useful in modelling claim data sets with many complex and overlapping features, e.g.,

- Taylor & McGuire (2004): Auto liability with rates of claim settlement varying over time, superimposed inflation varying with payment quarter and operational time, and legislative change causing a shift in claim cost according to accident quarter;
- Taylor & Mulquiney (2007): a mortgage insurance cascaded model with sub-models for healthy policies, in arrears, and properties in possession;

- Taylor, McGuire & Sullivan (2008); a medical malpractice model of individual claim development, using covariates such as specialty, geographic area of practice, etc.

The practical application of GLMs to reserving is described in a monograph by Taylor & McGuire (2016). Nowadays, GLM applications are regarded as a subset of **predictive modelling**.

3. The Paleogene: Increased Diversity in the Higher Forms

3.1. Adaptation of species

It may happen that model parameters shift over time, in which case (2.1) requires generalization to the following:

$$Y_{ij} = f\left(Y, \alpha^{(t)}\right) + \varepsilon_{ij}, E\left[\varepsilon_{ij}\right] = 0 \qquad (3.1)$$

where $\alpha^{(t)}$ is the parameter set α at time $t = i + j$, i.e. on the diagonal labelled $i + j$. The parameter set is viewed as evolving over time as $\alpha^{(t)} \sim P\left(.\,; \alpha^{(t-1)}\right)$, where P is some measure that depends on the previous parameter set $\alpha^{(t-1)}$.

Such models are called **evolutionary** or **adaptive**. They are reminiscent of the **Kalman filter** (Harvey, 1989) except that they usually involve stochastic errors other than the Gaussian ones required by Kalman. They have been investigated by Taylor (2008) and Taylor & McGuire (2009).

3.2. Miniaturization: parameter 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).

In this model, parameter estimation for model (3.1) is effected by minimization of the **penalized log-likelihood**

$$L(Y; \alpha) + \lambda \|\alpha\|_1 \qquad (3.2)$$

where $L(Y; \alpha)$ is the log-likelihood of the data set Y for parameter set α, $\|\alpha\|_1$ is the L_1-norm of α (i.e., the sum of the absolute values of all parameters), and $\lambda > 0$ is a discretionary parameter.

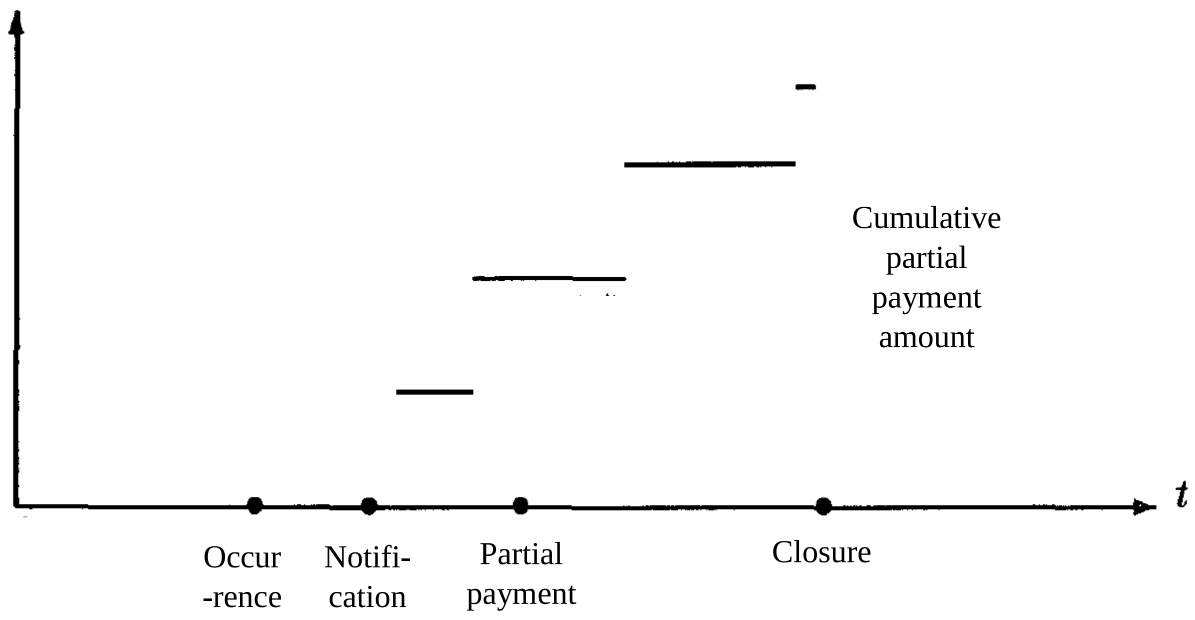
The second member of (3.2) is a penalty that increases as the number of model coefficients increases, and as their sizes increase. It serves to constrain model complexity. Indeed, a very useful property of the lasso is that it typically forces many components of α to zero

3.3. Granular (micro-) models

Granular models (also referred to as **micro-models**) analyse the fine detail of the claim process. Norberg (1993) represented this as a cascade of marked Poisson processes, as in Figure 3-1.

Figure 3-1

Norberg's Claim Process



These models are 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)).

A model of this sort would justify itself if it were to produce a loss reserve superior to that produced by aggregate methods. But recall that the (aggregate) chain ladder is minimum variance for ODP observations (Taylor, 2011), so a granular model would not enhance prediction efficiency in this case. Recall also that granular models will often consist of cascaded multiple sub-models, and the multiplicity is likely to inflate prediction error (Taylor & Xu, 2016).

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), so it is conditioned by more data than their aggregate models. This does not seem to compare like with like.

In summary, it appears that the jury is still out on the value of micro-models.

4. The Anthropocene Period: The Rise of Roboticus Sapiens

Recent years have seen the gradual introduction of **machine learning (ML)** into the loss reserving literature. A very early contribution came from Mulquiney (2006), who used an **artificial neural net (ANN)** to model a set of claim finalizations tabulated by accident quarter, development quarter, payment quarter, operational time at finalization and season (calendar quarter) of finalization.

The ANN proved effective in recognizing a number of known eccentricities in the data set. Specifically, it detected superimposed inflation that varied over both finalization quarter and operational time, and also the effects of a legislative change (accident quarter effect) that occurred in the midst of the claim experience. It produced goodness-of-fit superior to that of a parallel GLM.

More recently, ASTIN has established working parties to investigate the application of ML to loss reserving (Harej, Gächter & Jamal, 2017; Jamal *et. al*, 2018). The scope of the latter study extended beyond ANNs, also including **random forests** and **gradient boosting machines**. Other recent contributions have been made by Kuo (2018) and Wüthrich (2018).

5. The Later Anthropocene: The Watchmaker and The Soothsayer

The future appears to involve a contest between granular models and ML. The user of a granular model resembles a watchmaker, manipulating the microscopic components of the mechanism. The ANN user, on the other hand, does not specify a model at all, but trusts the neural net to digest all the detail of the data, make deductions from it, and speak forth a prophesy without particular explanation, as would a soothsayer.

As the volume of available claim data increases, there is a natural inclination to investigate the detail of the claim process in greater depth. This may be justified when the data are demonstrably at variance with the simpler forms of model.

A granular model will be a valuable aid to understanding of the claim process if it can be demonstrated effective in explaining the data. However, there are difficulties in formulating such a model. Some of these are referred to in Section 3.3.

But a possibly even greater difficulty arises from dependencies within the model. Consider, for example, Figure 3-1. It is unlikely that the amounts of partial payments are stochastically independent. If the observed partial payments in respect of a Liability claim consist of a number of small payments, followed by a much larger one, it is likely that this last is made in essential satisfaction of the claim, and the probability of a further large payment becomes small.

Consider, as another example, Workers Compensation claims, in which claimants may repeatedly move in and out of "Incapacitated" and "Active" states. One might account for this by modelling separately the rates of recovery from incapacity in spells 1, 2, etc., of incapacity, and similarly different rates of relapse for different spells of activity. However, it is unlikely that, for example, the event of relapse in say spell 3 of activity would be independent of the prior history of the claim. Relapse is probably more likely if the claim duration is only three weeks, as compared with three years.

The modelling of dependencies of this sort is difficult and laborious, and some models will contain a vast multiplicity of possible dependencies requiring investigation. Yet the model might become of dubious reliability if they do not receive attention. For these reasons, the potential of granular models may be limited.

The properties of ML are almost diametrically opposed to those of granular models. An ANN, for example, uses mainly simple numerical operations to produce an abstract function that matches a data set well. This unstructured form of model often yields little, if any, understanding of the claim process (the **interpretability problem**), and care will be required in forecasting because a model that fits well to past data does not necessarily predict the future equally well.

On the other hand, because of the ANN's abstract nature, it is well able to accommodate dependencies of the sort discussed above. There is, in my view, considerable promise in ML, but its main challenges will be to develop model interpretability and robustness of forecasts. For the present, the lasso discussed in Section 3.2 is a relatively structured form of ML. Its model form is available to the user, and is, to a certain extent, interpretable.

Greg Taylor, Ph.D., FIA, FIAA, FIMA, is an adjunct professor in the School of Risk and Actuarial Studies at University of New South Wales in Sydney, Australia. He can be reached at gregory.taylor@unsw.edu.au.

References

Antonio, K. and R. Plat. 2014. Micro-Level Stochastic Loss Reserving for General Insurance. *Scandinavian Actuarial Journal*, 2014(7): 649–669.

Baxter L. A., S. M. Coutts and S.A.F. Ross. 1979. Applications of Linear Models in Motor Insurance. *Proceedings of the 21st International Congress of Actuaries*, vol. 2, no. 11.

Benktander, G. 1976. An Approach to Credibility in Calculating IBNR for Casualty Excess Insurance. *The Actuarial Review*, 3, no. 7.

Bornhuetter R.L. and R.E. Ferguson. 1972. The Actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, vol. 59, no. 112, pp. 181–195.

Brockman M. J. and T. S. Wright. 1992. Statistical Motor Rating: Making Effective Use of Your Data. *Journal of the Institute of Actuaries*, 119, no. 3, 457–526.

Bühlmann, H. 1983. Estimation of IBNR Reserves by the Methods of Chain Ladder, Cape Cod and Complimentary Loss Ratio. International Summer School, unpublished.

Bühlmann, H. 2016. Historical Origin of the Cape Cod Claims Reserving Method. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2752387.

Fisher, W. H. and J. T. Lange. 1973. Loss Reserving Testing: a Report Year Approach. *Proceedings of the Casualty Actuarial Society*, vol. 60, 189–207.

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 Congress of Actuaries*, 1, 185–194.

Hachemeister, C. A. and J.N. Stanard 1975. IBNR Claims Count Estimation with Static Lag Functions. Spring Meeting of the Casualty Actuarial Society.

Harej B., R. Gächter and S. Jamal 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.

Hesselager, O. 1994. A Markov Model for Loss Reserving. *Astin Bulletin*, 24(2), 183-193.

Hovinen, E. 1981. Additive and Continuous IBNR. ASTIN Colloquium, Loen, Norway.

Huang, J., X. Wu and X. Zhou. 2016. Asymptotic Behaviors of Stochastic Reserving: Aggregate versus Individual Models. *European Journal of Operational Research*, 249, 657–666.

Jamal, S., S. Canto, R. Fernwood, C. Giancaterino, M. Hiabu, L. Invernizzi, T. Korzhynska, Z. Martin and H. Shen 2018. Machine Learning & Traditional Methods Synergy in Non-Life Reserving. Report of the ASTIN Working Party of the International Actuarial Association. Available at https://www.actuaries.org/IAA/Documents/ASTIN/ASTIN_MLTMS%20Report_SJAMAL.pdf.

Kuo, K. 2018. Deep Triangle: A Deep Learning Approach to Loss Reserving. **arXiv**:1804.09253v1 [stat.AP] Apr. 24, 2018.

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.

Nelder, J. A. and R. W. M. Wedderburn. 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.

Pigeon, M., K. Antonio and M. Denuit. 2013. Individual Loss Reserving with the Multivariate Skew Normal Framework. *Astin Bulletin*, 43(3), 399–428.

Pigeon, M., K. Antonio and M. Denuit. 2014. Individual Loss Reserving Using Paid-Incurred Data. *Insurance: Mathematics and Economics*, 58, 121–131.

Reid, D. H. 1978. Claim Reserves in General Insurance. *Journal of the Institute of Actuaries*, 105, 211–296.

Taylor, G. 2000. *Loss Reserving: An Actuarial Perspective*. Dordrecht, Netherlands: Kluwer Academic Publishers.

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. and M. Campbell. 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.

Taylor, G. and G. McGuire. 2004. Loss Reserving with GLMs: A Case Study. Casualty Actuarial Society 2004 Discussion Paper Program, 327–392.

Taylor, G. and G. McGuire. 2009. Adaptive Reserving Using Bayesian Revision for the Exponential Dispersion Family. *Variance*, 3, 105–130.

Taylor, G. and G. McGuire. 2016. Stochastic Loss Reserving Using Generalized Linear Models. CAS Monograph Series, Number 3. Appears at https://www.casact.org/pubs/monographs/index.cfm?fa=taylor-monograph03.

Taylor, G., McGuire, G. and J. Sullivan. 2008. Individual Claim Loss Reserving Conditioned by Case Estimates. *Annals of Actuarial Science*, 3(1–2), 215–256.

Taylor, G. and P. Mulquiney. 2007. Modelling Mortgage Insurance as a Multi-State Process. *Variance*, 1, 81–102.

Taylor, G. and J. Xu. 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, 579–604.

Wright, T. S. 1990. A Stochastic Method for Claims Reserving in General Insurance. *Journal of the Institute of Actuaries*, 117, 677–731.

Wüthrich, M. V. 2018. Neural Networks Applied to Chain-Ladder Reserving. Appears at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2966126.