



Annual Review of Statistics and Its Application

Recent Challenges in Actuarial Science

Paul Embrechts and Mario V. Wüthrich

RiskLab, Department of Mathematics, ETH Zurich, Zurich, Switzerland, CH-8092;
email: paul.embrechts@math.ethz.ch, mario.wuethrich@math.ethz.ch

Annu. Rev. Stat. Appl. 2022. 9:1.1–1.22

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

<https://doi.org/10.1146/annurev-statistics-040120-030244>

Copyright © 2022 by Annual Reviews.
All rights reserved

Keywords

actuarial science, generalized linear models, life and non-life insurance, neural networks, risk management, telematics data

Abstract

For centuries, mathematicians and, later, statisticians, have found natural research and employment opportunities in the realm of insurance. By definition, insurance offers financial cover against unforeseen events that involve an important component of randomness, and consequently, probability theory and mathematical statistics enter insurance modeling in a fundamental way. In recent years, a data deluge, coupled with ever-advancing information technology and the birth of data science, has revolutionized or is about to revolutionize most areas of actuarial science as well as insurance practice. We discuss parts of this evolution and, in the case of non-life insurance, show how a combination of classical tools from statistics, such as generalized linear models and, e.g., neural networks contribute to better understanding and analysis of actuarial data. We further review areas of actuarial science where the cross fertilization between stochastics and insurance holds promise for both sides. Of course, the vastness of the field of insurance limits our choice of topics; we mainly focus on topics closer to our main areas of research.

1. INTRODUCTION

Early pioneers in insurance mathematics were the Dutch statesman Johan de Witt (1625–1672), with his essay “The Worth of Life Annuities to Redemption Bonds,” and the Swiss theologian and mathematician Jakob Bernoulli (1655–1705), with his masterpiece *Ars Conjectandi*, where he proved an early version of the law of large numbers. In his correspondence with Gottfried Wilhelm Leibniz, Jakob Bernoulli mentioned in the context of his new asymptotic theory that the most important part of his work was still missing—namely, the application of his theoretical results to real world problems. Leibniz was not too enthusiastic about Bernoulli’s idea and argued that Bernoulli’s model was much too simple to answer real-world questions. It was his nephew Niklaus Bernoulli who later applied his uncle’s theory to mortality computations; we refer readers to Bolthausen & Wüthrich (2013).

From the very beginning, actuarial science has been a discipline in statistical modeling that has been driven by practical problems in insurance. Teaching rigorous mathematical lectures on actuarial science only gained ground much later. The theoretical cornerstones of actuarial science were the work of Cramér (1930, 1994) and the book by Bühlmann (1970), who introduced the stochastic model approach toward non-life insurance. The latter was in contrast with the more deterministic view of life insurance at that time. In an influential 1989 editorial in the *ASTIN Bulletin*, the editor then, Bühlmann (1989), famously introduced the so-called Actuary of the Third Kind, an actuary who uses his/her technical skills not only on the liability side of the insurance company’s balance sheet but also on the asset side. This actuary stands between the First Kind (the deterministic model-guided life actuary), the Second Kind (the stochastic model-oriented non-life actuary), and the Fourth Kind (the enterprise risk management-oriented actuary). These different kinds should not be interpreted as a reinvention of the actuarial profession; rather, they reflect the evolving societal conditions (demographic, technological, environmental, political, and legal) within which the actuarial profession fulfills its important tasks. On several occasions, we have defined the Actuary of the Fifth (final!) Kind as a data driven and model guided, critical and socially responsible financial decision maker in an ever changing world governed by uncertainty. As such, the Fifth Kind is not all that distant from the etymology of the word *actuarius*, originating in the mid-sixteenth century as meaning copyist or account keeper. In Roman times, the *actuarius* was a quartermaster keeping the legion’s books—so, surely, someone strongly linked with and helpful in reaching business or strategic decisions based on data.

Lester (2004) discusses the major challenges facing the actuarial profession going forward. They include the following: (a) the world has become a more uncertain place, (b) there are numerous vigorous competitors, (c) the corporate governance and transparency push is placing increasing responsibility on boards and senior management, (d) communication is becoming a key success factor, and (e) the actuarial vocation is growing and spreading. Though written in 2004, points a–e not only remain true but have become more accentuated. Modern society no doubt offers many new challenges for actuaries. Examples include supply chain insurance, crop insurance, longevity bonds, the evolving world of catastrophe insurance, pandemic bonds, parametric insurance, and innovative pension systems in a historically low-interest-rate environment, as well as the always-present market for insurance linked securities (ILSs). Areas like personalized medicine and telematics for auto insurance are newly born, drones take to the sky, and robots replace humans at an increasing rate. Several (but surely not all) of these new developments are driven by so-called big data and data science. The only viable way forward for actuaries is to embrace data science techniques or work closely with data scientists. A key anchor point, however, remains that an insurance product by (legal) definition offers a policyholder the relief of losses due to risks encountered. These products have to be well defined, technically understood, properly

marketed, regulatory agreed on, to a certain extent not discriminating, and correctly priced and reserved, as well as clearly communicated. This will always call for actuarial understanding and an education that goes beyond mathematics and statistics. A modern actuarial qualification includes education in legislation, economics, accounting, professional behavior, and communication, and such a qualification also asks for a recognized program of continuous professional development to not lose sight of the evolving state of the art.

In this review we mainly focus on recent challenges of statistical modeling in actuarial science. To this purpose, we divide actuarial science into three different branches: In Section 2, we discuss challenges in non-life insurance modeling; in Section 3, we present recent developments in statistical modeling of life insurance; and in Section 4, we study reinsurance, risk management and specific applications to operation risk and cyber risk. This classification in three branches is quite typical—on the one hand, there are legal reasons for this partition of insurance because products in these three different branches typically require different legal entities for their marketing and sale. On the other hand, these branches have rather different characteristics (risk drivers) that require different statistical modeling approaches. A bit outstanding is health and accident insurance, which has features of both life and non-life insurance as well as a strong intersection with social insurance, the latter being organized quite differently from the others.

The confines of a relatively short article on a field as vast as actuarial science limit not only the topics we can treat but also the depth we can go into for those topics we discuss. We very much hope, however, that the more statistically oriented reader will get a good feeling for the kind of statistical techniques that enter the field of actuarial science. The rather extensive list of references offers sources for further reading.

2. NON-LIFE INSURANCE MODELING

2.1. A Brief Overview of Non-life Insurance

The term non-life insurance is mainly used in Europe, and it summarizes all direct insurance products that are different from life insurance. In the United Kingdom, non-life insurance is also termed general insurance, and in the United States and Canada it is called property and casualty insurance. In non-life insurance, one typically distinguishes two functions: There is the pricing actuary who designs and prices (new) insurance products, and there is the reserving actuary who predicts cash flows of insurance claims (which typically run over multiple years). These predictions are used for insurance accounting, risk management, and product development.

2.2. Non-life Insurance Pricing

Non-life insurance pricing is the actuarial domain that runs at the forefront of statistical modeling and data science. It is a traditional discipline where the statistical modeling cycle is explored; here we refer readers to McCullagh & Nelder (1983, section 2.1) and Box & Jenkins (1976). Typically in non-life insurance, one faces the problem of having a heterogeneous portfolio of insurance policyholders, and one aims at charging risk-adjusted prices to each of these customers. This is a classical regression problem where one tries to find the systematic effects in data that discriminate policyholders. In contrast to many other fields in statistical modeling, actuarial problems are not based on causal graphs (and relations) but, at best, suitable proxies are found that explain propensity to claims. Actuaries are mainly interested in best predictions, however, at a level of model complexity for which products and prices can be clearly communicated to management and customers. Here we are reminded of the discussion of Breiman (2001), which lies at the heart of the conflict of choosing the best predictive algorithm versus the requirement of explainability.

2.2.1. From generalized linear models to neural networks. Most of the actuarial pricing methods used today are based on the seminal work of Nelder & Wedderburn (1972) and McCullagh & Nelder (1983) on generalized linear models (GLMs). Currently, actuarial pricing in car insurance is typically based on 40 to 50 covariates that discriminate policyholders. Over the past decades, the actuarial profession has gained much practical experience in engineering this information to make it useful for predictive modeling within GLMs. Actuaries have to cope with a couple of noteworthy difficulties. First of all, the majority of explanatory variables are of categorical type, and as a consequence, a statistical analysis faces complications such as, e.g., the sparsity of the underlying design matrix. Furthermore, in the regression functions, covariates interact in a nontrivial way, making proper estimation a challenging task. Claims frequency modeling is a rare event prediction problem (class imbalance problem) where actuaries try to find systematic effects in data that are heavily dominated by the noisy (random) part. Because there is no simple off-the-shelf distributional model, claim size modeling aims at finding a good compromise between model complexity and accuracy. This, for instance, holds true within the exponential dispersion family (EDF) that usually does not suit the whole range of claim sizes. For this reason, increasingly complex models are widely explored with resulting technical complications. For instance, there is an active stream of literature on mixture models (see Lee & Lin 2010, Miljkovic & Grün 2016, Yin & Lin 2016, Fung et al. 2019).

The ever-growing repository of data, however, makes it increasingly difficult to maintain manual feature engineering,¹ and actuaries have to rely more and more on representation learning² using tools like neural networks. There are many active actuarial communities that drive this field of research; we may, for instance, refer to the actuarial data science initiative of the Swiss Association of Actuaries (<https://www.actuarialdatascience.org/>) that produces tutorials and computer code based on publicly available insurance data.³

2.2.2. The balance property. GLMs are based on the EDF. If we use GLMs with the canonical link then they enjoy a particularly nice property, which in actuarial science is called the balance property. This property appears in the original work of Nelder & Wedderburn (1972). We briefly describe this balance property because it plays a crucial role in actuarial and financial modeling.

Assume we have a sequence of independent random variables Y_1, \dots, Y_n that can be described by the following one-dimensional linear EDF density:

$$Y_i \sim f(y; \theta_i, v_i/\varphi) = \exp \left\{ \frac{y\theta_i - \kappa(\theta_i)}{\varphi/v_i} + a(y; v_i/\varphi) \right\}, \quad 1.$$

where

- $v_i > 0$ is the given exposure (weight, volume) of risk $1 \leq i \leq n$;
- $\varphi > 0$ is the dispersion parameter;
- $\theta_i \in \Theta$ is the canonical parameter of risk $1 \leq i \leq n$ in the effective domain Θ ;
- $\kappa : \Theta \rightarrow \mathbb{R}$ is the cumulant function; and
- $a(\cdot; \cdot)$ is the normalization, not depending on the canonical parameter θ .

¹We denote by feature engineering the process of extracting features (covariates) from raw data using (insurance) domain knowledge to select the relevant information.

²We denote by representation learning the process of letting machine learning tools explore suitable covariate structure from raw data so that it can be used in a regression analysis.

³A nice source of publicly available insurance data is the R Core Team (2018) package `CASdatasets` of Dutang & Charpentier (2019).

Expression 1 defines a density with respect to a given σ -finite measure on \mathbb{R} ; for more details on the EDF, we refer readers to Barndorff-Nielsen (2014) and Jørgensen (1986, 1987, 1997). This model enjoys the properties

$$\mu_i = \mathbb{E}_{\theta_i} [Y_i] = \kappa'(\theta_i) \quad \text{and} \quad \text{Var}_{\theta_i} (Y_i) = \frac{\varphi}{v_i} \kappa''(\theta_i).$$

For a GLM we choose a suitable link function g such that we can express the systematic effects as follows:

$$\mathbf{x}_i \mapsto g(\mu_i) = g(\kappa'(\theta_i)) = \langle \boldsymbol{\beta}, \mathbf{x}_i \rangle, \quad 2.$$

where $\boldsymbol{\beta} \in \mathbb{R}^{q+1}$ is the regression parameter, $\mathbf{x}_i \in \{1\} \times \mathbb{R}^q$ is the covariate information, and $\langle \cdot, \cdot \rangle$ is the scalar product in the Euclidean space \mathbb{R}^{q+1} . Importantly, the covariate information $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,q})^\top$ includes an intercept term (the constant) in the first component.

If we choose the canonical link $g = (\kappa')^{-1}$, the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\beta}}^{\text{MLE}}$ of $\boldsymbol{\beta}$ enjoys the balance property on the considered portfolio—that is,

$$\sum_{i=1}^n v_i \widehat{\mu}_i = \sum_{i=1}^n v_i \mathbb{E}_{\widehat{\theta}_i} [Y_i] = \sum_{i=1}^n v_i \kappa' \left(\widehat{\boldsymbol{\beta}}^{\text{MLE}}, \mathbf{x}_i \right) \stackrel{!!}{=} \sum_{i=1}^n v_i Y_i. \quad 3.$$

This means that the total premium $\sum_{i=1}^n v_i \widehat{\mu}_i$ charged over the entire portfolio (for the next accounting year) equals the total claim amount $\sum_{i=1}^n v_i Y_i$ observed in the last accounting year. Why is this important? Suppose that we do not change the underlying portfolio. In that case the resulting pricing principle is unbiased—in fact, it is uniformly minimum variance unbiased. This is crucial in actuarial and financial applications because it explains that the price charged over the entire portfolio is on the right level: Underpricing would lead to financial ruin in the short or long run, whereas if we overpriced the portfolio, we would lose our market competitiveness. However, the balance property in Equation 3 is even more remarkable. Note that it holds no matter whether the chosen GLM is suitable or not. That is, even under a completely wrong GLM for describing the observations Y_1, \dots, Y_n , the balance property holds. This means that the prices $v_i \widehat{\mu}_i$ may be completely misspecified on an individual policy level i , but nevertheless, we charge the right overall price $\sum_{i=1}^n v_i \widehat{\mu}_i$ over the entire portfolio $1 \leq i \leq n$. Of course, one should try to avoid this latter kind of unintended cross-subsidy because it would also imply that we are not competitive on certain parts of the portfolio.

This example of the balance property shows that there are required key properties in actuarial modeling that differ from typical applications in statistics. On the one hand, large parts of actuarial modeling rely on MLE; however, unbiasedness is a key property in actuarial and financial pricing. On the other hand, we emphasize the importance of MLE or, equivalently, of minimizing the corresponding deviance loss function. We mention this because in many applications it is crucial to choose a problem adapted objective function for model calibration. This requires an extended domain knowledge about the data and the problem at hand (which links our discussion to the modeling cycle mentioned at the beginning of Section 2.2): Data collection, data cleaning, and data preprocessing are key to the understanding and selection of a problem-adapted modeling solution. In low-frequency problems and potentially large event simulations, the models need to be able to appropriately judge such observations, which is hardly the case with the squared-error loss function that is predominantly used in recent data science developments. Robustness is another important property that the procedures should have. However, outliers in insurance typically are not data errors but large financial claims that are an important pricing component; note that insurance is a domain where often the largest 20% of claims account for 80% of the total claim amount (we refer readers to the lecture notes of Wüthrich 2020b, figure 3.5). This corresponds to the so-called 20-80 Pareto rule (see, for instance, Embrechts et al. 1997, section 8.2.3).

2.2.3. Neural networks and representation learning. Above, we have emphasized the role of the actuary and his/her domain knowledge for preprocessing data to make it suitable for GLMs. Increasingly, the task of data preprocessing is taken over by machine learning methods. On the one hand, this is motivated by the fact that networks may find structure beyond actuarial domain knowledge, and on the other hand, increasing data repositories make it difficult to cope with the speed of new data collection. In this sense, we would like to see (feed-forward) neural networks as an extension of GLMs. This extension is most easily understood by revisiting the regression function in Expression 2 as follows:

$$\mathbf{x}_i \mapsto g(\mu_i) = g(\kappa'(\theta_i)) = \langle \boldsymbol{\beta}, \mathbf{z}^{(d:1)}(\mathbf{x}_i) \rangle, \tag{4}$$

where $\mathbf{z}^{(d:1)}$ is a composition of $d \in \mathbb{N}$ hidden neural network layers $\mathbf{z}^{(d:1)} = \mathbf{z}^{(d)} \circ \dots \circ \mathbf{z}^{(1)}$ (we use the symbol \circ for compositions of functions). These hidden neural network layers $\mathbf{z}^{(k)} : \mathbb{R}^{q_{k-1}} \rightarrow \mathbb{R}^{q_k}$ are nonlinear transformations of raw covariate information so that the learned representations $\mathbf{z}_i = \mathbf{z}^{(d:1)}(\mathbf{x}_i)$ are suitable inputs for the GLM in (Expression 4). An example of depth $d = 3$ is illustrated in **Figure 1a**. The three hidden layers (black circles) preprocess the covariate information $\mathbf{x}_i \mapsto \mathbf{z}_i = \mathbf{z}^{(3:1)}(\mathbf{x}_i)$, and the orange box gives the GLM in Expression 4 on these learned representations \mathbf{z}_i .

Based on well-known universality theorems (see, e.g., Cybenko 1989, Hornik et al. 1989, Isenbeck & Rüschemdorf 1992), we know that large enough feed-forward neural networks are sufficiently flexible to approximate any continuous and compactly supported regression function arbitrarily well. Moreover, Elbrächter et al. (2021) have recently proven that deep networks enjoy better convergence properties over shallow ones. This emphasizes that we should use sufficiently large deep feed-forward neural networks for feature engineering in Expression 4. State-of-the-art network calibration uses variants of the stochastic gradient descent (SGD)

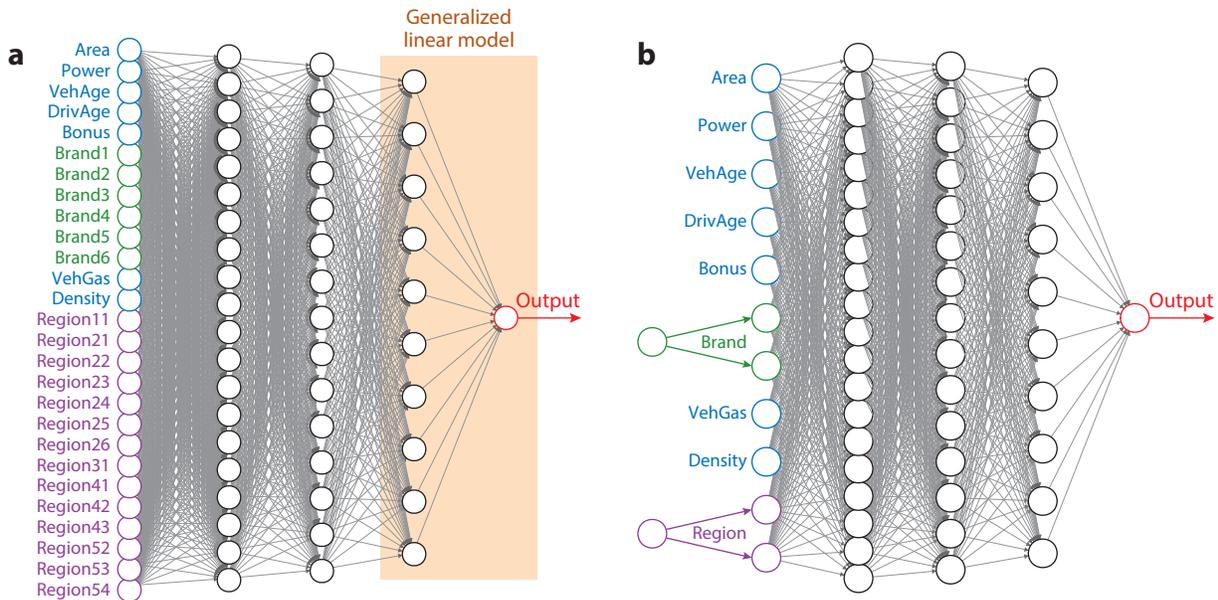


Figure 1 (a) Feed-forward neural network of Expression 4 of depth $d = 3$. (b) Feed-forward neural network using embedding layers of dimension 2 for categorical covariates (green and purple).

algorithm (see Goodfellow et al. 2016). Early stopping of these SGD algorithms is a crucial regularization technique to prevent these large networks from overfitting to the (in-sample) data. This early stopping plays an essential role in successfully fitting the neural network regression model; however, it also induces some undesirable side effects:

- First, early stopping implies that the fitted model does not correspond to a critical point of the objective function. This implies that, under the choice of the deviance loss function, an early stopped neural network calibration cannot be an MLE and, henceforth, the balance property similar to Equation 3 will typically fail. This has been pointed out by Wüthrich (2020a), who discusses methods to eliminate this deficiency.
- Second, each run of the SGD algorithm requires a starting value (seed), and early stopping implies that for each different starting value we typically receive a different neural network calibration and, henceforth, different prices for the same insurance policyholder. This is particularly troublesome in insurance because it implies that insurance prices have an element of randomness resulting in different prices for the same customer in different runs of the algorithm. Emphasizing the importance that insurance prices need to be explainable to management and customers, this implies that such methods cannot directly be used for insurance pricing (since this element of randomness may imply price fluctuations that cannot be objectively justified). There have been some attempts that try to average (blend) over multiple neural network prices (see, e.g., Richman & Wüthrich 2020). A final solution to this problem has not yet been found since maintaining multiple neural network models is not feasible economically. In fact, the claims frequency problem studied by Richman & Wüthrich (2020) needs averaging over 400 different network calibrations to obtain suitable stability for prices on an individual policy level.

Another interesting feature in neural network modeling is embedding layers that allow for a different treatment of categorical covariates from dummy coding in GLMs (see Richman 2021a,b). These recent developments are strongly inspired by natural language processing (see Bengio et al. 2003, 2013). These authors propose to embed categorical variables into low-dimensional Euclidean spaces such that proximity in embeddings is equivalent to similarity for regression modeling. Such embeddings typically reduce the complexity of the network: In **Figure 1**, we illustrate the same regression problem in a situation where we have two categorical covariates. The network in **Figure 1a** uses dummy coding for these two covariates, and the network in **Figure 1b** uses two-dimensional embedding layers. Obviously, these embedding layers for categorical covariates reduce the complexity of the network. Optimal embeddings of categorical variables can also be learned as part of SGD model calibration. Such learned embeddings allow for useful graphical illustrations, possibly augmented by a principal component analysis (PCA) and/or a K -means clustering step. Taken together, they contribute toward an efficient representation and statistical explanation of the chosen models. An example is found in Perla et al. (2021, figure 8) for a US mortality study using the US states' information as categorical variables. The resulting learned representation has surprising similarity with the US map, indicating that neighboring states have similar mortality behavior (which, of course, makes perfect sense).

We conclude this subsection with a few remarks and recent developments.

- Increasingly, unstructured data are used in regression models. In insurance, unstructured data are mostly available in the form of claims reports, medical reports, etc. A first case study in this direction has been conducted by Lee et al. (2020).
- The neural network regression model in Expression 4 does not directly build on improving the weaknesses of a preliminarily chosen GLM, nor does it help to improve a GLM. A simple way to build on an existing GLM is to boost this GLM with a neural network (or

any other machine learning approach); this approach has been emphasized in the *ASTIN Bulletin* editorial by Wüthrich & Merz (2019).

- Neural networks are often criticized for not being interpretable because feature engineering and variable selection are done internally by the neural network in a nontransparent way. In a recent work, Richman & Wüthrich (2021) introduce the so-called LocalGLMnet, which is a new neural network architecture that shares many properties of GLMs. This new architecture provides an additive decomposition that is interpretable, allows for variable selection, and enables interaction effects to be studied.
- At the moment, neural network modeling is still at the early stage of finding good algorithms for best predictors. As emphasized by Hans Bühlmann in a panel discussion at the Insurance Data Science Conference 2019 (<https://insurancedatascience.org/>), we should only start to call a predictive method scientific once we are able to quantify prediction uncertainty. Unfortunately, many fields of modern machine learning are not there yet.
- The above GLM and neural network approaches are sometimes also called a priori pricing approaches because they rely on policyholder information that is available at execution of the contract. For renewals of contracts, individual past claims experience of policyholders is available, which allows the insurer to include experience rating in renewal prices. An active field of actuarial research is the design of bonus-malus systems (BMSs) that rewards good claims experience with bonuses. BMSs go back to, and have been developed in, the work of Loimaranta (1972), De Pril (1987), Lemaire (1995), Denuit et al. (2007), Brouhns et al. (2003), and Ágoston & Gyetvai (2020). This stream of literature mainly studies optimal designs of BMSs with their economic implications, such as the bonus hunger of customers. A second stream of literature studies instead the question of optimally extracting feature information from an existing BMS to predict future claims (see Boucher & Inoussa 2014, Verschuren 2021).

2.2.4. Discrimination and telematics car driving data. As described in the introduction to Section 2.2, available characteristics of insurance policyholders are used as proxies to explain propensity to claims. For instance, the age of a car driver might be a good proxy to describe driving experience and skills. This practice may lead to discussions with insurance customers because, for instance, not every young car driver is a bad driver. Moreover, by law, there are protected characteristics that are not allowed to be used for insurance pricing. For instance, in Europe, gender is a so-called protected variable that is not allowed to influence insurance prices. The use of other characteristics may not (yet) be forbidden by law but is treated as problematic from a reputational point of view; one such example is ethnicity. Legislation on discrimination has recently raised many discussions in the actuarial community, and as a consequence, more and more contributions in the actuarial literature focus on such questions; Guillén (2012) and Chen et al. (2018) provide two references on this topic. A key issue in developing so-called discrimination-free prices is to ensure that discrimination also does not take place indirectly, because the more information we have about policyholders, the better we can determine their protected features. This has been studied by Lindholm et al. (2020), who give a proposal that avoids direct and indirect discrimination. An important finding of their proposal is that fully avoiding (indirect) discrimination requires full knowledge of protected features of customers. Of course, the latter might be rather problematic in practice because the insurer has to have full information about gender, ethnicity, and so on in order to mitigate indirect discrimination (on these covariates), otherwise he/she is not able to unravel all the different effects. Apparently, this discrimination-free pricing framework

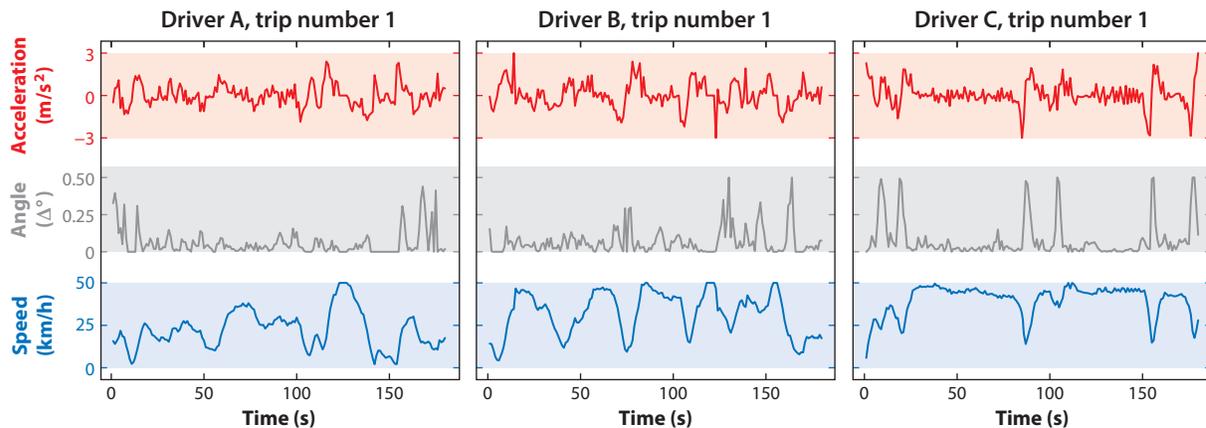


Figure 2

Individual trip of drivers A, B, and C. The top curve (*red*) shows acceleration/braking, the middle one (*gray*) shows the change in direction over 180 s, and the bottom curve (*blue*) shows the speed.

has strong similarity with the do-operator in causal statistics, and it is also closely related to the partial dependence plots of Friedman (2001) and Zhao & Hastie (2021).

There is some hope that the discussion of discrimination can be circumvented by personal and behavioral data, such as telematics car driving data. Such data are much more targeted toward propensity to claims, but their use also raises legal issues, this time related to privacy concerns. Telematics car driving data are high-frequency data that record driving information, say, second by second. For instance, they may record Global Positioning System location, speed, acceleration, change of direction, etc. every second. Moreover, they may give information about road conditions, traffic, weather conditions, time stamps of trips, driving above speed limits, etc. Thus, such data are very transparent about driving habits and driving style of car drivers, and it seems rather clear that this will characterize propensity to claims much better than any other feature. The main issues in handling this data are the size of the data (typically in the magnitude of terabytes) and the precision of telematics data. Recent actuarial work has started to extract features from telematics data. On the one hand, specific information is extracted from these data such as speeding, hard acceleration, and total distances (see, e.g., Ayuso et al. 2016, 2019; Boucher et al. 2017; Huang & Meng 2019; Lemaire et al. 2016; Paefgen et al. 2014; Sun et al. 2020; Verbelen et al. 2018). On the other hand, a more integral approach is taken by Weidner et al. (2016) and Gao et al. (2019, 2021): These papers directly focus on extracting driving style features from telematics data, which can then be entered into regression models. The former paper uses a Fourier decomposition, whereas in the latter papers the telematics data are directly processed through a neural network.

We close this section with a small example taken from Gao & Wüthrich (2019) that highlights the privacy concerns that should be raised when using telematics data. We select at random three car drivers from our portfolio and study their speed, acceleration, and change of angle patterns of their individual trips. **Figure 2** shows for each of these three selected drivers (called A, B and C) a car driving trip of length 180 seconds. Gao & Wüthrich (2019) showed that it is not too difficult to train a convolutional neural network classification model to correctly allocate such patterns of (only) 180 seconds to the right driver. Thus, from telematics data, it is not too difficult to determine which family member is currently driving the car, whether the car driver is in good health, and so on. Not surprisingly, society is alarmed by such transparency, and policymakers

need to clearly define the conditions under which what kind of data can be used to guarantee a necessary level of privacy.

2.3. Claims Reserving in Non-life Insurance

Claims reserving in non-life insurance is concerned with predicting claims cash flows that can last over multiple years; an introduction to claims reserving is provided by the monograph by Wüthrich & Merz (2008). Claims reserving can roughly be divided into four different historical stages. The first stage was the algorithmic time, when actuaries developed algorithms to predict claims cash flows. Only in the second stage were these algorithms lifted to full stochastic models; we mention here the path-breaking work of Mack (1993) and Renshaw & Verrall (1998). The uncertainty tools developed at that time focused on a static total uncertainty view considering simultaneously the entire lifetime of the cash flows. This static view is not in line with modern solvency and accounting requirements where incoming new information constantly changes best predictors and, therefore, prediction uncertainty has to be understood as a dynamic process. In the third stage emerged the notion of claims development results that can be viewed as martingale innovations (see Merz & Wüthrich 2008, 2014; Röhr 2016). These first three stages focused on modeling aggregated claims, and only very recently have developments on individual claims reserving started to flourish in the fourth stage—of course, benefiting heavily from data science tools such as regression trees and neural networks. The stochastic basis for this individual claims reserving view was already introduced in the 1990s by Arjas (1989) and Norberg (1993, 1999). First real data applications of these stochastic concepts have still been too rigid in practice (see Antonio & Plat 2014, Pigeon et al. 2013). However, recently, considerable progress has been made in this area of actuarial modeling; references are Baudry & Robert (2019), Delong et al. (2021), Gabrielli (2020), Kuo (2020), Lopez et al. (2019), Wang et al. (2021) and Wüthrich (2018).

One difficulty in this field of research is the availability of data. Usually, individual claims reserving data are confidential, protecting the privacy of injured people and business strategies of insurance companies. This makes it difficult to further develop these tools. There are recent initiatives that aim at generating synthetic individual claims data that share the same features as real data. Currently, two stochastic scenario generators are available (see Gabrielli & Wüthrich 2018, Avanzi et al. 2020).

3. LIFE INSURANCE MODELING

3.1. A Brief Overview of Life and Pension Insurance

The nature of life and pension insurance is rather different from non-life insurance; the latter typically offers financial protection against unforeseen events over the next accounting year. Life and pension insurance insures life and protects against disability and death of individual policyholders over possibly their entire lifetimes. For instance, one can buy an annuity product that guarantees fixed payments over the entire remaining lifetime of the policyholder. Such a product can be bought by a single up-front premium payment at the inception of this multiple-year contract. As a consequence, life insurance is very much concerned with predicting mortality and longevity trends over several decades into the future. This prediction is typically based on time series models. Furthermore, life insurance needs to organize investments and hedging of long-term financial guarantees that are granted at inception to the policyholders. This requires good financial and economic models, as well as suitable optimization tools for multi-period portfolio optimization.

3.2. Mortality Modeling

Clearly, the more statistical part of life and pension insurance modeling is mortality forecasting. Mortality forecasting has a long tradition. As mentioned in the first paragraph of the Introduction, by the seventeenth century, governments had started to organize social old-age welfare systems that were based on mortality projections. Today, the most popular stochastic mortality projection models are the Lee & Carter (1992) (LC) model and the Cairns et al. (2006) (CBD, for Cairns, Black & Dowd) model. Since then, a vast literature on stochastic mortality modeling has developed, with most of the approaches being offspring and generalizations of either the LC or the CBD model. A taxonomy of the most commonly used models is provided by Cairns et al. (2009, table 1). This progress accelerated after the turn of the millennium and the subsequent financial crisis. The financial crisis has led to a low-interest-rate environment where more accurate mortality projections have gained crucial importance. In a low-interest-rate environment, misspecification of longevity trends can no longer be covered by high financial returns on investments. Moreover, national social security systems have also recently come under financial pressure, which makes the field of mortality projection a central object of interest to politicians, economists, and demographers [see, e.g., Hyndman & Ullah (2007) and Hyndman et al. (2013) for relevant literature in demography]. We briefly describe the LC model to mathematically ground this discussion on mortality and longevity projection. Typically, one starts with national mortality data obtained from the Human Mortality Database (HMD) (<https://www.mortality.org>). These data comprise two components: the death counts $d_{t,x}$ and the exposures $e_{t,x}$ by calendar year t and age x . These data are collected separately for both genders and over several countries in the HMD. Based on this information, one calculates the raw (crude) mortality rate for age x in calendar year t by

$$m_{t,x} = d_{t,x}/e_{t,x}.$$

Since age x can be defined in different ways for a given calendar year t and since the exposures $e_{t,x}$ can be distorted by migration (i.e., raw mortality rates are not calculated on a closed population), data quality has to be analyzed carefully at this stage.

In **Figure 3**, we illustrate raw log-mortality rates $\log(m_{t,x})$ for calendar years 1900–2016 and ages 0–99 of Swiss females and Swiss males. We observe the familiar patterns: Female mortality is lower than male mortality, population mortality is decreasing over time, and major improvements in mortality are observed after World War II. Furthermore, in the male heat map, we see the emergence of HIV for males aged between 20 and 40 after 1980 (the light blue band). The Spanish flu in calendar years 1918–1920 leads to a visible vertical structure. Based on these observations, we aim at projecting mortality beyond the latest observed calendar year.

We now describe the LC model. The LC model can be understood as a regression model based on categorical variables for calendar year t and age x . It makes the following structural assumption:

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{t,x}, \quad 5.$$

where

a_x is the average force of mortality at age x ,

k_t is the time index describing the change of force of mortality for calendar year t ,

b_x is the rate of change of force of mortality broken down to the different ages x , and

$\varepsilon_{t,x}$ are independent and centered error terms.

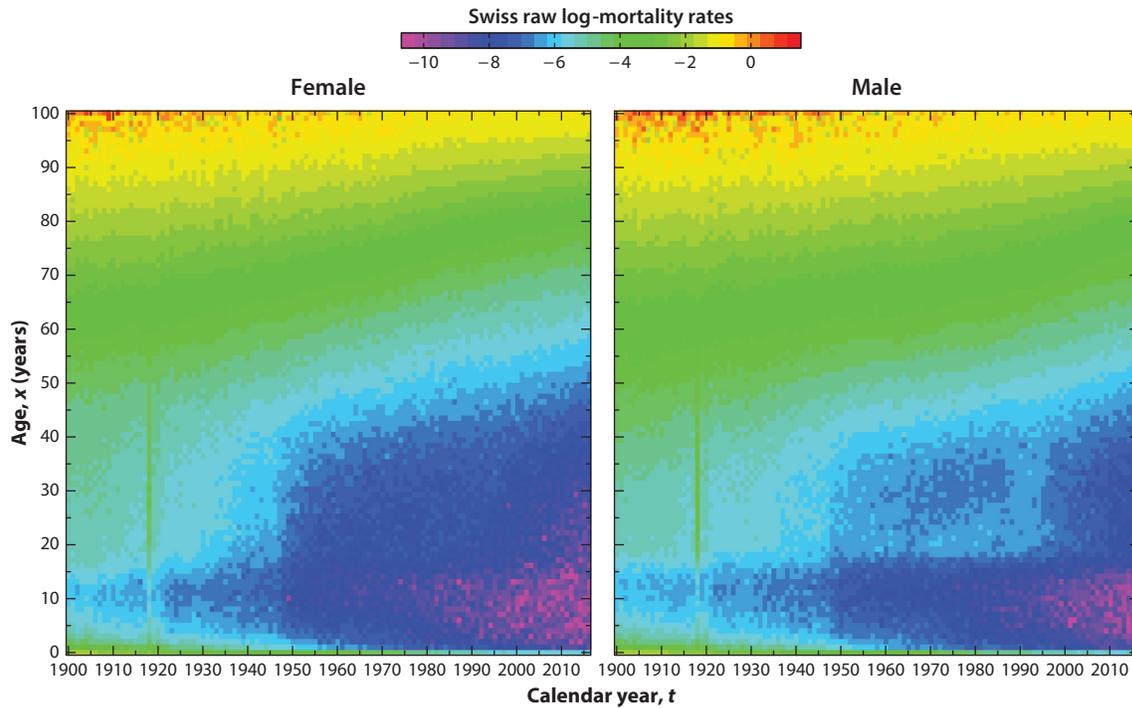


Figure 3

Raw log-mortality rates $\log(m_{t,x})$ for calendar years 1900–2016 and ages 0–99 of Swiss females (*left*) and Swiss males (*right*); both plots use the same color scale. The different colors illustrate different levels of mortality ranging from low mortality (*magenta-blue*) to high mortality (*yellow-orange*).

The LC model is fitted to each population and each gender separately—i.e., the model given in Equation 5 does not immediately allow us to consider multiple populations simultaneously. Fitting and prediction is done in two steps. In the first step, one fits the parameters $(a_x)_{x_0 \leq x \leq x_1}$, $(k_t)_{t_0 \leq t \leq t_1}$, and $(b_x)_{x_0 \leq x \leq x_1}$ to the available raw log-mortality data $\mathbf{M} = (\log(m_{t,x}))_{t_0 \leq t \leq t_1, x_0 \leq x \leq x_1}$. This estimation step can be done by using the singular value decomposition (SVD) providing the first principal component as a description of the centered log-mortality data matrix of \mathbf{M} . In the second step, the SVD estimated change of force of mortality $(\hat{k}_t)_{t_0 \leq t \leq t_1}$ is extrapolated beyond calendar year t_1 using a time series model, which in the simplest case is a random walk with drift.

The LC model of Equation 5 has been generalized in many directions. First, note that by using the SVD, we implicitly assume that the error terms in Equation 5 are Gaussian. Brouhns et al. (2002) modified model estimation to the more reasonable Poisson case. Renshaw & Haberman (2006) added a cohort effect $b_x^{(1)}\gamma_{t-x}$ to the LC model because they observed that in many populations there is a strong cohort effect in mortality data. Hyndman & Ullah (2007) explored a functional data method using penalized regression splines, and Hainaut & Denuit (2020) adopted this approach to a wavelet-based decomposition enjoying favorable properties in time-series modeling. Shang (2019) extended the classical LC model based on a static PCA to a dynamic PCA regression, and Villegas et al. (2018) studied computational aspects of model fitting.

Recent research has increasingly focused on joint mortality modeling of multiple populations: We mention the common age effect model of Kleinow (2015), the augmented common factor model of Li & Lee (2005), and the functional time series models of Hyndman et al. (2013) and

Shang & Haberman (2020). Not surprisingly, machine learning methods have also moved into this area of statistical modeling. For instance, Perla et al. (2021) gave a large-scale multiple population forecast that is based on recurrent neural networks and convolutional neural networks, respectively. This model allows for an interpretation within the LC structure.

Another interesting area of research is the study of long-memory in mortality data (Yan et al. 2020) and socioeconomic differences in mortality (Cairns et al. 2019). In relation to the current coronavirus 2019 (COVID-19) pandemic, one rather explores continuous time hazard rate models because these can easily be distorted on shorter timescales. We expect these models to be researched in more detail in the near future, also including investigations on cause of mortality and healthcare device tracking.

3.3. Insurance Product Design and Valuation of Cash Flows

Because life and pension insurance are of a long-term nature, product design is especially important to guaranteeing long term financial stability. This involves cash flow valuation, hedging of long-term financial guarantees, and minimizing of lifetime ruin probabilities. There is a vastly growing literature in this field of actuarial science that is based on stochastic process modeling, optimal control, and, increasingly, on machine learning methods like neural networks or reinforcement learning. It would go too far at this stage to dive into the actuarial literature on these topics; therefore, we only give selected interesting aspects. Clearly, it is crucial to have good stochastic models that allow us to project cash flows into the future and value these cash flows for insurance pricing, accounting, and risk management. For cash flow valuation one typically separates mortality risk drivers from financial and economic risk drivers, more mathematically speaking, by assuming that these risk drivers can be described by independent stochastic processes. This independence assumption then allows different valuation methods to be implemented and calculated more easily. These are mostly based on the no-arbitrage principle resulting in the consideration of martingales for price processes. Relevant work in this field of research has been done by DeLong et al. (2019a,b) and Deelstra et al. (2020).

These valuation methods play a crucial role in solvency assessments, long-term investment considerations, and naturally, in product design. Related to the latter, we focus on one particular idea that has recently gained some popularity. In view of increasing mortality improvements, low-interest-rate environments, and increasing regulatory constraints, private life insurance companies are more and more reluctant to offer long-term longevity and financial guarantees to customers. Therefore, in the actuarial literature, the old idea of the so-called tontine has returned. The first tontine system was developed in 1653 by the Italian Lorenzo de Tonti, and tontines gained much popularity between the seventeenth and nineteenth century, especially in France and the United Kingdom. In those times, tontines were used as investment plans to raise capital.

A tontine is a financial scheme that is organized by either a government or a corporation. Everyone can subscribe to a tontine scheme by paying an amount of capital into the tontine. This investment then entitles the subscriber to receive an annual interest until he/she dies. When a subscriber of the tontine scheme passes away, his/her share is reallocated among the survivors of this subscriber. This process terminates as soon as the last subscriber has died. Thus, the tontine essentially is a self-organizing pension system that does not involve an insurance company, and it also does not involve any longevity guarantees. It only needs a body that organizes the scheme and that manages the capital of the subscribers. In contrast to private (personal) investments, family members will not inherit tontine shares in case of death of the tontine subscriber, but these shares go instead to the survivors in the tontine scheme.

Such tontine considerations have recently gained popularity in the actuarial literature. Sabin (2010) studies the fair tontine annuity where subscribers can join the scheme at any time; they may have different age and gender and, thus, different expected remaining lifetimes. Under these conditions, Sabin (2010) defines a fair scheme, proves under which assumptions such a scheme exists, and provides an algorithm for constructing such a fair transfer plan. This stream of life and pension insurance has been carried forward by Milevsky & Salisbury (2015); Chen et al. (2020) study optimal combinations between tontine shares and classical annuity life insurance, and Bernhardt & Donnelly (2021) study minimal sizes of such annuity funds to achieve income stability. Naturally, an accurate mortality table is crucial in the construction of a fair transfer plan according to Sabin (2010) because tontine subscribers may have different age and gender to which changes of mortality can act rather differently. Note that such tontines do not have a sponsor in terms of an insurance company that covers longevity risk, but the tontine scheme has to manage and organize change of mortality itself. Similar ideas have recently also come up in non-life insurance under the name peer-to-peer insurance (see Denuit 2020).

4. RISK MANAGEMENT AND SELECTED TOPICS

4.1. Reinsurance

By definition, the insurance industry offers risk mitigation products and tools for its customers. The industry itself functions under carefully worked out risk management guidelines in order to be able to fulfill its contractual requirements toward these policyholders as well as to be attractive to its investors. The reinsurance industry provides specific services and products so that the primary (direct) insurers, referred to as ceding companies, can cap or reengineer their insurance product portfolios. Within the confines of this article, we are not able to enter into a detailed discussion of the reinsurance business, but we offer some reference pointers. In 2014 the Swiss Reinsurance Company Ltd, Swiss Re for short, celebrated its 150-year anniversary. On that occasion several publications described the industry (see, e.g., James et al. 2013, Haueter & Jones 2016). Swiss Re's sigma publications (<https://www.swissre.com/institute/research/sigma-research.html>) yield up-to-date developments of the industry and the underlying markets. A more mathematical introduction is given by Albrecher et al. (2017).

An important development that enters the reinsurance world relates to so-called alternative risk transfer (ART) products. These products typically originate in the realm of financial institutions—Wall Street, say—and become attractive especially when the reinsurance market is under stress because of high insured losses. Also driving ART markets are arguments of risk diversification; actuarial losses are perceived as having low correlation to financial markets (for an overview on ART, see, e.g., Culp 2002). In this context we also like to mention ILSs as investment instruments whose value is driven by insurance loss events (see, e.g., Barrieu & Albertini 2009). An excellent business barometer for the ART and ILS markets, as well as useful repository for interesting publications, is the website of Morton Lane, one of the pioneers in the field (<http://lanefinancialllc.com/>). It is no coincidence that the ART market for CAT products experienced a peak after the occurrence of Hurricane Andrew in August of 1992. Early analyses of the Chicago Board of Trade's 1992 futures written on US property insurance contracts were provided by Embrechts & Meister (1997) and Cummins & Geman (1995). These markets continue to pose major methodological challenges for actuaries.

The reinsurance industry, especially, faces serious challenges emerging from climate change due to its perceived impact on climatological catastrophes and related insurance covers. There currently is an explosion of publications, both academic and industry based, on this topic. Some industry-based examples are presented in a publication from the Zurich Insurance Group

(2019). Other noteworthy publications include that of Bevere & Gloor (2020) and the excellent position paper of the CRO Forum (2019). We leave it to the reader to sift through the almost daily appearing academic papers on climate change. We highlight some issues that no doubt play a dominant role from an actuarial point of view. First, climate change typically spreads over longer time periods, and hence discounting over such time spans is important. In this context, the work of Gollier (2001, 2013) is relevant. Second, the climate change debate is marred by considerable uncertainty about predictions. Here it is worthwhile to look at the so-called Dismal Theorem of Weitzman (2009, 2011), one of the fathers of the economics of climate change. In the face of deep structural uncertainty, Weitzman questioned classical cost-benefit analysis. This led to intensive debates, especially with Nordhaus, laureate of the 2018 Nobel Prize in Economic Sciences (see, e.g., Nordhaus 2009, 2011). Of course, in insurance, the problem with (non-)insurability in markets under deep structural uncertainty, very fat tails, is well known (see, e.g., Ibragimov et al. 2015). Such markets often lead to diversification disasters (Ibragimov et al. 2011) as well as to nondiversification traps (Ibragimov et al. 2009). The mathematical background to some of the above results in insurance and economics is provided by Embrechts et al. (1997) and McNeil et al. (2015).

4.2. From Risk Measures to Capital Adequacy

Since the early 1970s, capital adequacy regulation for banking and insurance has undergone major changes. The thrust of these changes clearly has been from a more traditional rules-based solvency system to a principal-based one leading to forward-looking risk-based guidelines. Different governments and industries implement these rules in different ways: For banking, these are the subsequent Basel guidelines (currently in transition from Basel III to IV); for European insurers, they are the different solvency guidelines (currently Solvency II); and for Switzerland, insurances are regulated via the Swiss Solvency Test. Risk-based capital rules require insurance companies to hold capital in relation to their risk profiles to guarantee that they have enough financial resources to withstand financial difficulties. McNeil et al. (2015, chapter 1) give an overview and further references; in chapter 2, they review the main developments concerning risk measures. A risk measure maps a risk position (e.g., a financial instrument, a portfolio of insurance contracts, a book of loans) to a real number, where McNeil et al. (2015) interpret a positive value as a net-risk position for which regulatory capital has to be put aside. There exists a considerable literature on the various sets of mathematical axioms financial and insurance risk measures have to satisfy; we refer readers to Föllmer & Schied (2011) for an in-depth discussion and further references.

Mainly driven by international regulatory capital requirements (Basel and Solvency), two risk measures came to the forefront: first, the quantile-type value-at-risk (VaR), and second, expected shortfall (ES), also referred to as tail VaR or conditional VaR. VaR originated in the early 1990s through the RiskMetrics publications from J.P. Morgan (see, e.g., Jorion 2007). Early on, academics warned that VaR in general fails to be subadditive. We note that VaR, as a quantile risk measure, answers questions of the if type and is frequency oriented. In contrast, ES is of the more important what-if type and is subadditive, indeed coherent (see Acerbi & Tasche 2002). Theorem 8.28 of McNeil et al. (2015) is key to understanding the link between the model assumptions on a portfolio $\mathbf{X} = (X_1, \dots, X_n)$ of one-period risk factors and the resulting properties of VaR-type risk measures. The key assumption in the theorem is that \mathbf{X} is elliptically distributed (see McNeil et al. 2015, definition 6.25). In this case, risk management is standard: VaR is subadditive; stress scenarios can be based on elliptical stress regions; and for a wide class of risk measures (including VaR and ES), Markowitz mean-variance portfolio optimization yields the same optimal portfolio. The theorem also yields methodological justification for the standard formula (solvency capital requirement) under Solvency II for risk aggregation [see McNeil et al.

2015, section 8.4, and the European Insurance and Occupational Pensions Authority (EIOPA) (<https://www.eiopa.europa.eu>]). Dhaene et al. (2006) give a review of risk measurement in insurance.

Once a risk measure has been defined for a given risk management problem, the next steps involve statistical estimation, goodness of fit (backtesting) given historical data, and practical implementation and communication. On these topics, very many scientific papers, books, and industry internal as well as regulatory documents have been written. We mention McNeil & Frey (2000) linking extreme value theory (EVT) with dynamic VaR estimation and backtesting. In order to compare the forecasting properties of competing risk measures, the work on elicibility is relevant; Gneiting (2011) provides a general paper. Applications to finance and insurance are discussed by Ziegel (2016), Nolde & Ziegel (2017) and Davis (2017). A further interesting paper on backtesting is that of Gordy & McNeil (2020). In this context, we definitely want to highlight Davis (2016). The overall area remains a highly relevant and very active field of research with important practical applications.

4.3. From Operational Risk to Cyber Risk

Every year, the insurance company Allianz publishes its Risk Barometer listing the top corporate perils (<https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/Allianz-Risk-Barometer-2020.pdf>). The 2020 version is based on the insight of more than 2,700 risk management experts from 102 countries and territories. The annual corporate risk survey is conducted among Allianz customers (global businesses), brokers and industry trade organizations. The top three risks are (a) cyber incidents (e.g., cyber crime, IT failure/outage, data breaches, fines and penalties); (b) business interruption (including supply chain disruption); and (c) changes in legislation and regulation. These risks are akin to operational risks. Under the EIOPA, operational risk is defined as “the risk of loss arising from inadequate or failed internal processes, personnel or systems, or from external events. Operational risks include compliance/legal risks, but exclude reputational, strategic and political/regulatory risks” (EIOPA 2019, p. 2). Operational risk no doubt constitutes a fundamental risk class for banks as well as for insurance companies. This became clear in the aftermath of the financial crisis, when large international banks lost a substantial percentage of market capitalization mainly due to legal fines linked to toxic financial instruments sold in the run-up to the crisis. Though cyber risk data have rather distinctive features, from a statistical point of view, they nonetheless show several communalities with operational risk data: nonhomogeneity, heavy tailedness, and intricate interdependencies between the various subcategories. Consequently, statistical modeling and estimation are difficult. Here, actuaries have a lot to offer. Operational risk data, and for that matter cyber risk data, show many properties well known to non-life actuaries. Textbook references on operational risk include Cruz et al. (2015) and McNeil et al. (2015, chapter 13) on insurance analytics and operational risk. The story behind the name “insurance analytics” is told by Embrechts (2002); the story stresses the fact that tools from actuarial mathematics are canonical for the modeling of operational and cyber risk. Because of the extreme heavy tailedness of operational risk data, EVT methodology is relevant (see, e.g., Chavez-Demoulin et al. 2016, Embrechts et al. 2018). Nešlehová et al. (2006) discuss the problem of extreme heavy-tailedness; readers are also directed to the references therein. There is a direct link to the Dismal Theorem of Weitzman (2009, 2011), discussed in Section 4.1 above.

The need for interdisciplinary approaches for handling cyber risk is discussed by Falco et al. (2019). Capital requirements for cyber risk and cyber risk insurance are discussed by Eling & Schnell (2020). An interesting overview on insurance and cyber risk in Sweden is that of Franke (2017); the paper also contains an extensive list of references on cyber-insurance in general.

4.4. A Comment on COVID-19

The Allianz Risk Barometer included, as major risks facing industry going forward, supply chain and business interruption risk. Both (obviously related) risks have been accentuated by the COVID-19 pandemic. Actuaries are to play a key role in these areas. Regulators early on warned about the threat facing society from viral pandemics and the potential disruption they may cause on supply chains. Business interruption risk was singled out as an especially important concern (see, e.g., Tripart. Auth. 2008). An important pre-COVID-19 example of supply chain risk is the 2011 Thai flood (see Gale & Saunders 2013). Also, the emergence of COVID-19 will have important consequences to life and health insurance, both on the statistical modeling and on the insurance product sides.

5. CONCLUSION

In our definition of the Actuary of the Fifth Kind in Section 1, we stressed the fact that an actuary is working in an ever changing world governed by randomness. Whereas our article gives several examples of this, one could have replaced “randomness” by “randomness and technology.” We have seen that data science is becoming increasingly ubiquitous for the modern actuary. Beyond the more statistically oriented development, changes are also taking place more at the technical core of information technology; these changes will have an impact not only on products but also on how insurance markets of the future and the various agents in those markets function. One example is that insurance companies have started to store their data in the cloud, where huge computational resources are available and where many insurance processes can be evaluated in real time. Of course, the latter is very much at the core of the discussion of whether to build statistical models that lead to a deeper understanding of structures in the data, or whether to be satisfied by just using big data to find the relevant correlations to make optimal predictions. We have emphasized that mathematics and statistics will likely play a crucial role in actuarial modeling also in the future. Another example is the potential impact of blockchain-like technology as well as artificial intelligence. The interested reader can for instance consult Meeusen & Sorniotti (2017) for a first impression. In a recent issue, the *Economist* published an article titled “Ping An. Metamorphosis. The world’s most valuable insurer has transformed itself into a fintech super-app. Could others follow its lead?” Time no doubt will tell. This brings us to a conclusion concerning the actuarial profession. In 2003, the readers of the Institute and Faculty of Actuaries’ publication *The Actuary* voted Frank Mitchell Redington (1906–1984) as the greatest British actuary ever. His collected publications and speeches, posthumously published as Redington (1986), yield important insight into the development of twentieth-century actuarial thinking, especially in the life insurance and pensions industries. Perhaps his most famous quote is, “the actuary who is only an actuary is not an actuary” (Benjamin & Redington 1968, p. 348). This quote very much reflects the ever changing world and the way in which actuaries over the centuries have used their technical skills for the better of society. As we have seen, statistics does play a fundamental role in this respect!

DISCLOSURE STATEMENT

Paul Embrechts and Mario Wüthrich are fully qualified actuaries of the Swiss Association of Actuaries. Mario Wüthrich is editor-in-chief of *ASTIN Bulletin*.

ACKNOWLEDGMENTS

The authors would like to thank David Rayment, Librarian of the Institute and Faculty of Actuaries, for the reference to the quote by Redington in Section 5.

LITERATURE CITED

- Acerbi C, Tasche D. 2002. On the coherence of expected shortfall. *J. Bank. Finance* 26:1487–503
- Ágoston KC, Gyetvai M. 2020. Joint optimization of transition rules and the premium scale in a bonus-malus system. *ASTIN Bull.* 50:743–76
- Albrecher H, Beirlant J, Teugels JL. 2017. *Reinsurance: Actuarial and Statistical Aspects*. New York: Wiley
- Antonio K, Plat R. 2014. Micro-level stochastic loss reserving for general insurance. *Scand. Actuar. J.* 2014(7):649–69
- Arjas E. 1989. The claims reserving problem in non-life insurance: some structural ideas. *ASTIN Bull.* 19:139–52
- Avanzi B, Taylor G, Wang M, Wong B. 2021. SynthETIC: an individual insurance claim simulator with feature control. *Insur. Math. Econ.* 100:296–308
- Ayuso M, Guillén M, Nielsen JP. 2019. Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation* 46:735–52
- Ayuso M, Guillén M, Pérez-Marín AM. 2016. Using GPS data to analyse the distance traveled to the first accident at fault in pay-as-you-drive insurance. *Transp. Res. Part C* 68:160–67
- Barndorff-Nielsen O. 2014. *Information and Exponential Families: In Statistical Theory*. New York: Wiley
- Barrieu P, Albertini L. 2009. *The Handbook of Insurance-Linked Securities*. New York: Wiley
- Baudry M, Robert CY. 2019. A machine learning approach for individual claims reserving in insurance. *Appl. Stoch. Model. Bus. Ind.* 35:1127–55
- Bengio Y, Courville A, Vincent P. 2013. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal.* 35:1798–828
- Bengio Y, Ducharme R, Vincent P, Jauvin C. 2003. A neural probabilistic language model. *J. Mach. Learn. Res.* 3:1137–55
- Benjamin B, Redington FM. 1968. Presentation of Institute Gold Medal to Mr Frank Mitchell Redington. *J. Inst. Actuar.* 94:345–48
- Bernhardt T, Donnelly C. 2021. Quantifying the trade-off between income stability and the number of members in a pooled annuity fund. *ASTIN Bull.* 51:101–30
- Bevere L, Gloor M. 2020. *Natural catastrophes in times of economic accumulation and climate change*. Rep. Sigma 2, Swiss Re Inst., Zurich, Switz.
- Bolthausen E, Wüthrich MV. 2013. Bernoulli's law of large numbers. *ASTIN Bull.* 43:73–79
- Boucher J-P, Côté S, Guillén M. 2017. Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks* 5(4):54
- Boucher J-P, Inoussa R. 2014. A posteriori ratemaking with panel data. *ASTIN Bull.* 44:587–12
- Box GEP, Jenkins GM. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day
- Breiman L. 2001. Statistical modeling: the two cultures. *Stat. Sci.* 16:199–15
- Brouhns N, Denuit M, Vermunt JK. 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insur. Math. Econ.* 31:373–93
- Brouhns N, Guillén M, Denuit M, Pinquet J. 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. *J. Risk Insur.* 70:577–99
- Bühlmann H. 1970. *Mathematical Methods in Risk Theory*. New York: Springer
- Bühlmann H. 1989. Editorial: actuaries of the Third Kind? *ASTIN Bull.* 19:5–6
- Cairns AJG, Blake D, Dowd K. 2006. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *J. Risk Insur.* 73:687–18
- Cairns AJG, Blake D, Dowd K, Coughlan GD, Epstein D, et al. 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *N. Am. Actuar. J.* 13:1–35
- Cairns AJG, Kallestrup-Lamb M, Rosenskjold C, Blake D, Dowd K. 2019. Modelling socio-economic differences in mortality using a new affluence index. *ASTIN Bull.* 49:555–90
- Chavez-Demoulin V, Embrechts P, Hofert M. 2016. An extreme value approach for modeling operational risk losses depending on covariates. *J. Risk Insur.* 83:735–76
- Chen A, Guillén M, Vigna E. 2018. Solvency requirement in a unisex mortality model. *ASTIN Bull.* 48:1219–43
- Chen A, Rach M, Sehner T. 2020. On the optimal combination of annuities and tontines. *ASTIN Bull.* 50:95–129

- Cramér H. 1930. *On the Mathematical Theory of Risk*. Stockholm: Centraltryckeriet
- Cramér H. 1994. *Collected Works*, Vols. I & II. New York: Springer
- CRO (Chief Risk Off.) Forum. 2019. *The heat is on—insurability and resilience in a changing climate*. Position Pap., CRO Forum, Amstelveen, Neth.
- Cruz MG, Peters GW, Shevchenko PV. 2015. *Fundamental Aspects of Operational Risk and Insurance Analytics*. New York: Wiley
- Culp CL. 2002. *The ART of Risk Management: Alternative Risk Transfer, Capital Structure, and the Convergence of Insurance and Capital Markets*. New York: Wiley
- Cummins D, Geman H. 1995. Pricing catastrophe insurance futures and call spreads: an arbitrage approach. *J. Fixed Income* 4:46–57
- Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* 2:303–14
- Davis MHA. 2016. Verification of internal risk measure estimates. *Stat. Risk Model.* 33:67–93
- Davis MHA. 2017. Discussion of "Elicitability and backtesting: perspectives for banking regulation." *Ann. Appl. Stat.* 11:1886–7
- De Pril N. 1978. The efficiency of a bonus-malus system. *ASTIN Bull.* 10:59–72
- Deelstra G, Devolder P, Gnameho K, Hieber P. 2020. Valuation of hybrid financial and actuarial products in life insurance by a novel three-step method. *ASTIN Bull.* 50:709–42
- Delong L, Dhaene J, Barigou K. 2019a. Fair valuation of insurance liability cash-flow streams in continuous time: applications. *ASTIN Bull.* 49:299–33
- Delong L, Dhaene J, Barigou K. 2019b. Fair valuation of insurance liability cash-flow streams in continuous time: theory. *Insur. Math. Econ.* 88:196–08
- Delong L, Lindholm M, Wüthrich MV. 2021. Collective reserving using individual claims data. *Scand. Actuar. J.* <https://doi.org/10.1080/03461238.2021.1921836>
- Denuit M. 2020. Investing in your own and peers' risks: the simple analytics of P2P insurance. *Eur. Actuar. J.* 10:335–59
- Denuit M, Maréchal X, Pitrebois S, Walhin J-F. 2007. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. New York: Wiley
- Dhaene J, Vanduffel S, Goovaerts MJ, Kaas R, Tang Q, Vyncke D. 2006. Risk measures and comonotonicity: a review. *Stoch. Models* 22:573–606
- Dutang C, Charpentier A. 2019. *CASdatasets R package vignette*. Ref. Manual, Version 1.0–10. <http://cas.uqam.ca/>
- Economist*. 2020. Ping An. Metamorphosis. The world's most valuable insurer has transformed itself into a fintech super-app. Could others follow its lead? *Economist*, Dec. 5–11, pp. 61–62
- EIOPA (Eur. Insur. Occup. Pension Auth.). 2019. *Opinion on the supervision of the management of operational risks faced by IORPs*. Work. Pap. EIOPA-BoS-19-247, EIOPA, Frankfurt am Main, Ger.
- Elbrächter D, Perekrestenko D, Grohs P, Böleskei H. 2021. Deep neural network approximation theory. *IEEE Trans. Inform. Theory* 67(5):2581–623
- Eling M, Schnell W. 2020. Capital requirements for cyber risk and cyber risk insurance: an analysis of Solvency II, the US Risk-Based Capital Standards, and the Swiss Solvency Test. *N. Am. Actuar. J.* 24:370–92
- Embrechts P. 2002. Insurance analytics. *Br. Actuar. J.* 8:639–41
- Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer
- Embrechts P, Meister S. 1997. Pricing insurance derivatives, the case of CAT-futures. In *Proceedings of the 1995 Bowles Symposium on Securitization of Risk*, pp. 15–26. Schaumburg, IL: Soc. Actuar.
- Embrechts P, Mizgier KJ, Chen X. 2018. Modeling operational risk depending on covariates: an empirical investigation. *J. Oper. Risk* 13:17–46
- Falco G, Eling M, et al. 2019. Cyber risk research impeded by disciplinary barriers. *Science* 6469:1066–69
- Föllmer H, Schied A. 2011. *Stochastic Finance: An Introduction in Discrete Time*. Berlin: Walter de Gruyter. 4th ed.
- Franke U. 2017. The cyber insurance market in Sweden. *Comput. Secur.* 68:130–44
- Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29:1189–32

- Fung TC, Badescu AL, Lin XS. 2019. A class of mixture of experts models for general insurance: application to correlated claim frequencies. *ASTIN Bull.* 49:647–88
- Gabrielli A. 2020. A neural network boosted double overdispersed Poisson claims reserving model. *ASTIN Bull.* 50:25–60
- Gabrielli A, Wüthrich MV. 2018. An individual claims history simulation machine. *Risks* 6(2):29
- Gale EL, Saunders MA. 2013. The 2011 Thailand flood: climate causes and return periods. *Weather* 68:226–38
- Gao G, Wang H, Wüthrich MV. 2021. Boosting Poisson regression models with telematics car driving data. *Mach. Learn.* <https://doi.org/10.1007/s10994-021-05957-0>
- Gao G, Wüthrich MV. 2019. Convolutional neural network classification of telematics car driving data. *Risks* 7(1):6
- Gao G, Wüthrich MV, Yang H. 2019. Driving risk evaluation based on telematics data. *Insur. Math. Econ.* 88:108–19
- Gneiting T. 2011. Making and evaluating point forecast. *J. Am. Stat. Assoc.* 106(494):746–62
- Gollier C. 2001. *The Economics of Risk and Time*. Cambridge, MA: MIT Press
- Gollier C. 2013. *Pricing the Planet's Future: The Economics of Discounting in an Uncertain World*. Princeton, NJ: Princeton Univ. Press
- Goodfellow I, Bengio Y, Courville A. 2016. *Deep Learning*. Cambridge, MA: MIT Press
- Gordy MB, McNeil AJ. 2020. Spectral backtests of forecast distributions with application to risk management. *J. Bank. Finance* 116:105817
- Guillén M. 2012. Sexless and beautiful data: from quantity to quality. *Ann. Actuar. Sci.* 6:231–34
- Hainaut D, Denuit M. 2020. Wavelet-based feature extraction for mortality projection. *ASTIN Bull.* 50:675–707
- Haueter NV, Jones G. 2016. *Managing Risk in Reinsurance: From City Fires to Global Warming*. Oxford, UK: Oxford Univ. Press
- Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. *Neural Netw.* 2:359–66
- Huang Y, Meng S. 2019. Automobile insurance classification ratemaking based on telematics driving data. *Decis. Support Syst.* 127:113156
- Hyndman RJ, Booth H, Yasmee F. 2013. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* 50:261–83
- Hyndman RJ, Ullah MS. 2007. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* 51:4942–56
- Ibragimov M, Ibragimov R, Walden J. 2015. *Heavy-Tailed Distributions and Robustness in Economics and Finance*. New York: Springer
- Ibragimov R, Jaffee D, Walden J. 2009. Nondiversification traps in catastrophe insurance markets. *Rev. Financ. Stud.* 22:959–99
- Ibragimov R, Jaffee D, Walden J. 2011. Diversification disasters. *J. Financ. Econ.* 99:333–48
- Isenbeck M, Rüschenhoff L. 1992. Completeness in location families. *Probab. Math. Stat.* 13:321–43
- James H, Borscheid P, Gugerli D, Straumann T. 2013. *Value of Risk: Swiss Re and the History of Reinsurance*. Oxford, UK: Oxford Univ. Press
- Jørgensen B. 1986. Some properties of exponential dispersion models. *Scand. J. Stat.* 13:187–97
- Jørgensen B. 1987. Exponential dispersion models. *J. R. Stat. Soc. Ser. B* 49:127–45
- Jørgensen B. 1997. *The Theory of Dispersion Models*. London: Chapman & Hall
- Jorion P. 2007. *Value-at-Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill. 3rd ed.
- Kleinow T. 2015. A common age effect model for the mortality of multiple populations. *Insur. Math. Econ.* 63:147–52
- Kuo K. 2020. Individual claims forecasting with Bayesian mixture density networks. arXiv:2003.02453 [stat.AP]
- Lee GY, Manski S, Maiti T. 2020. Actuarial applications of word embedding models. *ASTIN Bull.* 50:1–24
- Lee RD, Carter LR. 1992. Modeling and forecasting US mortality. *J. Am. Stat. Assoc.* 87(419):659–71
- Lee SCK, Lin XS. 2010. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *N. Am. Actuar. J.* 14:107–30

- Lemaire J. 1995. *Bonus-Malus Systems in Automobile Insurance*. Amsterdam: Kluwer Acad.
- Lemaire J, Park SC, Wang K. 2016. The use of annual mileage as a rating variable. *ASTIN Bull.* 46:39–69
- Lester R. 2004. *Quo vadis actuarius?* Paper presented at IACA, PBSS and IAA Colloquium, Sydney, Aust., Oct. 31–Nov. 5
- Li N, Lee R. 2005. Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. *Demography* 42:575–94
- Lindholm M, Richman R, Tsanakas A, Wüthrich MV. 2020. *Discrimination-free insurance pricing*. Work. Pap., ETH Zurich, Zurich. <http://dx.doi.org/10.2139/ssrn.3520676>
- Loimaranta K. 1972. Some asymptotic properties of bonus systems. *ASTIN Bull.* 6:233–45
- Lopez O, Milhaud X, Thérond P-E. 2019. A tree-based algorithm adapted to microlevel reserving and long development claims. *ASTIN Bull.* 49:741–62
- Mack T. 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bull.* 23:213–25
- McCullagh P, Nelder JA. 1983. *Generalized Linear Models*. London: Chapman & Hall
- McNeil AJ, Frey R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empir. Finance* 7:271–300
- McNeil AJ, Frey R, Embrechts P. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton Univ. Press. 2nd ed.
- Meeusen P, Sornioti A. 2017. *Blockchain in re/insurance: technology with a purpose*. Presentation, Swiss Re Inst., Rüslikon, Switz., Nov. 7
- Merz M, Wüthrich MV. 2008. Modelling the claims development result for solvency purposes. *CAS E-Forum* 2008(Fall):542–68
- Merz M, Wüthrich MV. 2014. *Claims run-off uncertainty: the full picture*. Res. Pap. 14–69, Swiss Finance Inst., Geneva. <https://dx.doi.org/10.2139/ssrn.2524352>
- Milevsky MA, Salisbury TS. 2015. Optimal retirement income tontines. *Insur. Math. Econ.* 64:91–05
- Miljkovic T, Grün B. 2016. Modeling loss data using mixtures of distributions. *Insur. Math. Econ.* 70:387–96
- Nelder JA, Wedderburn RWM. 1972. Generalized linear models. *J. R. Stat. Soc. Ser. A* 135:370–84
- Nešlehová J, Embrechts P, Chavez-Demoulin V. 2006. Infinite mean models and the LDA for operational risk. *J. Oper. Risk* 1:3–25
- Nolde N, Ziegel JF. 2017. Elicitability and backtesting: perspectives for banking regulation, with discussion. *Ann. Appl. Stat.* 11:1833–74
- Norberg R. 1993. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bull.* 23:95–115
- Norberg R. 1999. Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bull.* 29:5–25
- Nordhaus WD. 2009. *An analysis of the Dismal Theorem*. Discuss. Pap. 1686, Cowles Found. Res. Econ., Yale Univ., New Haven, CT
- Nordhaus WD. 2011. The economics of tail events with an application to climate change. *Rev. Environ. Econ. Policy* 5:240–57
- Paefgen J, Staake T, Fleisch E. 2014. Multivariate exposure modeling of accident risk: insights from pay-as-you-drive insurance data. *Rev. Environ. Econ. Policy* 61:27–40
- Perla F, Richman R, Scognamiglio S, Wüthrich MV. 2021. Time-series forecasting of mortality rates using deep learning. *Scand. Actuar. J.* <https://doi.org/10.1080/03461238.2020.1867232>
- Pigeon M, Antonio K, Denuit M. 2013. Individual loss reserving with the multivariate skew normal framework. *ASTIN Bull.* 43:399–28
- R Core Team. 2018. R: A language and environment for statistical computing. *Statistical Software*, R Found. Stat. Comput., Vienna
- Redington F. 1986. *A Ramble Through the Actuarial Countryside*. London: Staple Inn
- Renshaw AE, Haberman S. 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insur. Math. Econ.* 38:556–70
- Renshaw AE, Verrall RJ. 1998. A stochastic model underlying the chain-ladder technique. *Br. Actuar. J.* 4:903–23
- Richman R. 2021a. AI in actuarial science—a review of recent advances—part 1. *Ann. Actuar. Sci.* In press. <https://doi.org/10.1017/S1748499520000238>

- Richman R. 2021b. AI in actuarial science—a review of recent advances—part 2. *Ann. Actuar. Sci.* In press. <https://doi.org/10.1017/S174849952000024X>
- Richman R, Wüthrich MV. 2020. The nagging predictor. *Risks* 8(3):83
- Richman R, Wüthrich MV. 2021. LocalGLMnet: interpretable deep learning for tabular data. arXiv:2107.11059 [cs.LG]
- Röhr A. 2016. Chain-ladder and error propagation. *ASTIN Bull.* 46:293–30
- Sabin MJ. 2010. Fair tontine annuity. *SSRN Electron. J.* <https://dx.doi.org/10.2139/ssrn.1579932>
- Shang HL. 2019. Dynamic principal component regression: application to age-specific mortality forecasting. *ASTIN Bull.* 49:619–45
- Shang HL, Haberman S. 2020. Forecasting multiple functional time series in a group structure: an application to mortality. *ASTIN Bull.* 50:357–79
- Sun S, Bi J, Guillén M, Pérez-Marín AM. 2020. Assessing driving risk using internet of vehicles data: an analysis based on generalized linear models. *Sensors* 20(9):2712
- Tripart. Auth. 2008. *Market wide pandemic exercise 2006 progress report*. Rep., Financ. Serv. Auth., H.M. Treas., Bank Engl., London
- Verbelen R, Antonio K, Claeskens G. 2018. Unraveling the predictive power of telematics data in car insurance pricing. *J. R. Stat. Soc. Ser. C* 67:1275–304
- Verschuren RM. 2021. Predictive claim scores for dynamic multi-product risk classification in insurance. *ASTIN Bull.* 51:1–25
- Villegas AM, Millossovich P, Kaishev VK. 2018. StMoMo: stochastic mortality modeling in R. *J. Stat. Softw.* 84:1–32
- Wang Z, Wu X, Qiu C. 2021. The impacts of individual information on loss reserving. *ASTIN Bull.* 51:303–47
- Weidner W, Transchel FWG, Weidner R. 2016. Classification of scale-sensitive telematic observables for risk individual pricing. *Eur. Actuar. J.* 6:13–24
- Weitzman M. 2009. On modeling and interpreting the economics of catastrophic climate change. *Rev. Econ. Stat.* 91:1–19
- Weitzman M. 2011. Fat-tailed uncertainty in the economics of catastrophic climate change. *Rev. Environ. Econ. Policy* 5:275–92
- Wüthrich MV. 2018. Machine learning in individual claims reserving. *Scand. Actuar. J.* 2018(6):465–80
- Wüthrich MV. 2020a. Bias regularization in neural network models for general insurance pricing. *Eur. Actuar. J.* 10:179–202
- Wüthrich MV. 2020b. *Non-life insurance: mathematics & statistics*. Work. Pap., RiskLab, ETH Zurich, Zurich. <https://dx.doi.org/10.2139/ssrn.2319328>
- Wüthrich MV, Merz M. 2008. *Stochastic Claims Reserving Methods in Insurance*. New York: Wiley
- Wüthrich MV, Merz M. 2019. Editorial: Yes, we CANN! *ASTIN Bull.* 49:1–3
- Yan H, Peters GW, Chan JSK. 2020. Multivariate long-memory cohort mortality models. *ASTIN Bull.* 50:223–63
- Yin C, Lin XS. 2016. Efficient estimation of Erlang mixtures using iSCAD penalty with insurance application. *ASTIN Bull.* 46:779–99
- Zhao Q, Hastie T. 2021. Causal interpretations of black-box models. *J. Bus. Econ. Stat.* 39(1):272–81
- Ziegel JF. 2016. Coherence and elicibility. *Math. Financ.* 26:901–18
- Zurich Insurance Group. 2019. *Managing the impacts of climate change: risk management responses*. White Pap., Zurich Insur. Group, Zurich