Pricing German Health Insurance Products with Only Few Insured Persons

Jens Piontkowski[1,*]

[1] Mathematisches Institut der Heinrich–Heine–Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

[*] Correspondence: Jens Piontkowski <Jens.Piontkowski@hhu.de>

Abstract

If a health insurance product has only few insured, its claims experience becomes very volatile and is therefore not reliable enough as the only source for repricing the product. Traditionally, a similar product with many insured is used as a reference. However, legislative changes and market forces have led to a fragmentation of products. As a result, such a reference product with many insured is often no longer available. Here we propose a statistical model that combines the data of several products with few insured to derive a common relative claim inflation as well as the expected claims of these products in the future, thus enabling stable pricing for these products. The model was designed so that the usual premium adjustment process is changed as little as possible, making it easy to use in practice.

Key words German health insurance, Pricing, Premium adjustment, Short term forecasting, Bayesian model

Statements and Declarations

Competing Interests The author works for an insurance group.

1 Introduction

Due to legislation [1, 2] and in order to stay competitive German health insurance companies introduced many new products in recent years. As a consequence these new products have fewer policyholders than before and therefore their claim experience is more volatile. A working group of the German Actuarial Association (DAV) published a paper [3] indicating how many insured are needed to get a reliable estimate of an age–normalized claim per person based on the work of Siegel [4]. The task of estimating the medical inflation as well is even more difficult because it is often of the same order of magnitude as the statistical fluctuations. Previously, the yearly increase of claims was often taken from a similar older product (a technic known in German as "Stütztarif"). Here, products are referred to as similar when they offer nearly the same benefits in the areas of outpatient, inpatient, and dental care [3, Section 3.2]. Nowadays, many of the insured of these older products have changed to newer products and the claim experience of these older products have become less stable themselves. Thus the task is now to determine the yearly claim increase not from one similar product, but from a collection of similar products.

The general idea to deal with this problem is to scale the claims experience of all these similar products to a common level, then combine the claims, derive a trend for these combined claims as usual, and finally scale this down to the individual products, see Milbrodt and Röhrs [5, Chapter 5 Tarifklassen]. While Milbrodt and Röhrs assume the scaling factor be known beforehand, the DAV working group suggested an ad hoc estimator for the scaling factor [3, Section 3.2]. Gottschalk and Lax [6] pointed out that this estimator induced a bias in the trend whenever the relative distribution of insured in these products changes significantly during this time period and suggested a new one.

This solved the problem for the requirements of premium calculation in practice. In the thesis of Käsgen [7] the ad hoc estimator was shown to be consistent under the usual assumptions. By example it was also shown that the estimator is close to unbiased, but an attempt to prove this — if true at all — failed since the estimator does not have a simple distribution. Considering that the whole process consists of three steps, scaling up with the estimator, extrapolation, and finally scaling down with the estimator, it seems almost impossible to fully understand the whole process statistically.

Here, we replace the entire process with a statistical model and obtain the scale factors and the yearly claim increase for each product in one step. We can also estimate the variance of all the parameters. This allows us to understand how reliable the scaling factors and the predictions for the future are, which is especially important for the products with few insured.

In this paper we will use the Bayesian approach to modeling. This not only allows us to use our a priori knowledge, it also gives us the full distribution of the parameters we are interested in. Furthermore, we can easily take into account the correlation between the claims of the insured in successive years due to long–term and chronic illnesses, see Appendix B, — a topic, which has been neglected in the German health insurance literature so far.

Acknowledgments

The author wishes to express gratitude to Christian Lax and Lars Schmitz for bringing this intriguing topic to his attention, with special thanks to Christian Lax for his valuable suggestions on enhancing this exposition. Additionally, the author is thankful to the Deutsche Krankenversicherung (DKV) for supplying the data.

2 Notation and Setup

The main business of German health insurance is comprehensive health insurance, which replaces the mandatory statutory health insurance. A comprehensive health insurance contract for an adult is designed to be lifelong by the government [8, §192—208]. On the one hand, health insurance policyholders face a financial disadvantage if they terminate their contract. The intention is to prevent adverse risk selection by discouraging healthier insured from switching to another insurer. On the other hand, the health insurance company waives its right to terminate the contract, allowing insured with chronic illnesses to continue receiving coverage. Since the development of medical expenses can only be reliably predicted in the short term, the legislator allows for premium adjustments. To ensure that the policyholder is not exploited, this adjustment process is strictly regulated by a governmental ordinance [9], particularly regarding the decision of whether a premium adjustment may be carried out at all.

According to this regulation, the premiums are calculated to remain constant for life for each insured, assuming there is no medical inflation after the following year. This is achieved by equating the present value of a constant future premium of an insured person with the present value of his or her expected future claims. For estimating future claims of each insured, it is assumed that the expected claims of an insured depend only on age and gender. Even for the new unisex products, the claims estimation is initially done separately by gender and then mixed in the gender ratio, allowing us to focus on one gender for this study, leaving only age dependency of the claims. For details that need to be considered in practice, see Milbrodt and Röhrs [5, Chapter 5] or Becker [10, Section 3.2].

Let us recall how expected future claims are estimated for a product with many insured. We use the usual notation: Age is denoted by x, the set of all ages for adult insured is denoted by X, typically starting at 20 and ending with the terminal age of the life table, which is also used in the premium calculation [9, Anlage 1]. A product (German "Tarif") is denoted by θ, and a year is denoted by t. We denote the current year as $t = 0$ and assume that we know the claims experience of past years $T = \{t_{\min}, \dots, -1\}$. In practice, $t_{\min}$ is often between -6 and -3; in the examples, we use -5.

The essential input for calculating the premium for a product θ is the expected claims per insured, $K_x^\theta(t)$, for all ages x in the following year $t = 1$. These are estimated on the basis of observed average claims, ${}^{\text{obs}}K_x^\theta(t)$, in the past. The number of insured per age and year may be small, so the observations may be strongly influenced by random fluctuations. Another problem is that there may be no insured and therefore no observations for a product at certain ages, especially at higher ages. However, the expected claims at the higher ages are needed to calculate the present value of future claims.

During the development of actuarial methods in German health insurance in the 1930s and 1940s [11, chapters A to C], the method of Rusam became established to solve this problem [12]. It is assumed that expected claims per year and age can be divided into an age–dependent and a time–dependent part:

$$K_x^\theta(t) = k_x^\theta \cdot G^\theta(t),$$

where k_x^θ is the age profile (in German "Kopfschadenprofil") and $G^\theta(t)$ is an age–normalized claim (in German "Grundkopfschaden"). Note that to make the parameters of the Rusam model identifiable one needs an additional assumption; it is customary to set $k_{x_0}^\theta = 1$ for a reference age x_0. The basic idea behind this model is that all ages are relatively affected by medical inflation to the same extent. This assumption is a good approximation for short time periods. In the long term age profiles change [13], or other models fit better [14, 15]. In practice, however, only this formula by Rusam is used, as the profile can be adjusted every few years during premium adjustments, and long–term considerations are not relevant. In fact, the use of this formula is even mandated when assessing whether a premium adjustment can be made for a product [9, Anlage 2]. Using a different model for premium calculation would risk the models diverging, and that despite foreseeable losses, no premium adjustment could be made.

The general references mentioned at the beginning explain methods for deriving an age profile. This involves combining information from many products. Since small insurance companies do not have enough data to derive their own age profiles, the German government organization BaFin collects data from all health insurance companies to derive age profiles and publishes them annually, e.g. [16]. Since we assume that we are examining products with few insured, we do not have enough data to derive the profiles ourselves and assume that the age profiles of all products under consideration are given.

Therefore, our task of estimating the future $K_x^\theta(1)$ for all x has now simplified to estimating the future $G^\theta(1)$. We start by defining the estimator for the past $G^\theta(t)$, again adhering to the legal regulations [9, Anlage 2]: Denote by $l_x^\theta(t)$ the number of insured individuals of age x in year t in product θ then we expect the following total claim for the product:

$$S^\theta(t) := \sum_{x \in X} l_x^\theta(t) K_x^\theta(t) = \sum_{x \in X} l_x^\theta(t) k_x^\theta G^\theta(t) = G^\theta(t) \sum_{x \in X} l_x^\theta(t) k_x^\theta.$$

Turning this around the observed age–normalized claim is defined for the observed total claim, ${}^{\text{obs}}S^\theta(t)$, as

$${}^{\text{obs}}G^\theta(t) := \frac{{}^{\text{obs}}S^\theta(t)}{\sum_{x \in X} l_x^\theta(t) k_x^\theta}.$$

Having estimates for the past $G^\theta(t)$, $t < 0$, it remains to extrapolate to the future $t = 1$.

In German health insurance, it is common practice to assume a linear development of the age–normalized claims:

$$G^\theta(t) = a + bt \quad \text{for suitable } a, b.$$

Following Behne [17, Section 2], Milbrodt and Röhrs [5, Section 5.13(c)] also discuss the obvious alternative of assuming an exponential development, which is the idea behind the inflation indices used in similar contexts. However, this is not used in practice in Germany. The technical reasons are that the results of linear and exponential extrapolation do not differ significantly if inflation remains close to the usual 3%. Additionally, linear extrapolation is more robust than exponential extrapolation when there are so few data points. However, the main reason is most likely that the governmental ordinance [9, §15 and Appendix 2B] strongly recommends a linear extrapolation based on observations of the last three years to assess whether premiums can be adjusted to avoid losses. While other statistical methods are permitted, they must be established when introducing a new product or for a very compelling reason and, in both cases, approved by the German government organization BaFin, which has been avoided so far. Since linear extrapolation is almost mandatory for this purpose and no problems have arisen in the past when using it to calculate premiums, insurance companies prefer to continue using it to calculate premiums.

The age–normalized claim ${}^{\text{obs}}G^\theta(t)$ can be viewed as a weighted average of the claims of the insured. We expect it to follow a normal distribution — by definition around the mean $G^\theta(t)$. In practice, its variance is assumed to be constant over time for a product with many insured. This leads to a simple linear regression model, i.e. we have

$$ {}^{\text{obs}}G^\theta(t) = a + bt + \epsilon, \qquad \epsilon \sim \text{normal}(0, \sigma^2). $$

The future $G^\theta(1)$ is estimated by the extrapolation to $t = 1$.

3 The Model for a Set of Similar Products

Now assume that instead of one product θ with many insured, we have a set Θ of products with few insured. In order to proceed these must be similar in some way. Inspired by the definition of a product class (in German "Tariffklasse") in [5, Section 5.8] we make the following

Definition 3.1 (similar products)**.** A set of products Θ with given age profiles is called *similar* products, if and only if their age–normalized claim size is proportional to each other over time, i.e. there exist $a^\theta \in \mathbb{R}^{>0}$, $\theta \in \Theta$ such that

$$ \frac{G^{\theta_1}(t)}{a^{\theta_1}} = \frac{G^{\theta_2}(t)}{a^{\theta_2}} \qquad \text{for all } t \text{ and all } \theta_1, \theta_2 \in \Theta. $$

The *(fictitious) base age–normalized claim* is

$$ G(t) := \frac{G^\theta(t)}{a^\theta} \qquad \text{independent of } \theta \in \Theta. $$

In practice, it is assumed that products are similar when they offer nearly the same benefits in the areas of outpatient, inpatient, and dental care, because then they should be equally affected by medical inflation.

The definition of a product class has a different purpose and differs in two important aspects: Firstly, it is not required that the proportionality factors are time independent basically because it is assumed that we know them already a priori, for example because Θ is a set of product with proportional reimbursement. Secondly, it is assumed that the proportionality holds with the same factor for each claim size per age. We do not need to require the later, because we assume the age profiles as given. In practice, however, the same age profile is often used for similar products, so that proportionality also holds with the same factor for each expected claim size per age.

As mentioned above Gottschalk and Lax [6] introduced a natural ad hoc estimator for the proportionality factors. With these factors, all the claim experience can be scaled to the fictitious base product. Then one can combine all the claims and proceed as before for a product with many insured. After extrapolating the fictitious base product one needs to undo the scaling to get back to the individual products.

Here, we will combine all of these steps into one model that estimates everything in a single step. Assuming a linear trend for the fictitious base age–normalized claims

$$ G(t) = a + bt, $$

it follows that

$$ G^\theta(t) = a^\theta G(t) = a^\theta(a + bt) \qquad \text{for all } \theta \in \Theta. $$

There is an ambiguity in the choice of the scaling factors a^θ, since they can all be scaled by the same factor without violating the equations in the definition of similar products. We resolve this by requiring $1 = G(0) = a$. In summary, we assume that the expected age–normalized claims follow

$$ G^\theta(t) = a^\theta(1 + bt) \qquad \text{for all } \theta \in \Theta. $$

When extrapolating a product with many insured it is often assumed that the observation error is homoscedastic, because the number of insured in the product is nearly constant over time. Here we need to be more careful, because in general products with more insured will have smaller observation errors, and we want them to contribute more strongly to the determination of the common relative claim increase b. In fact, as ${}^{\text{obs}}G^{\theta}(t)$ is a weighted average of $l^{\theta}(t)$ observations we expect it to be normally distributed with mean $G^{\theta}(t)$ and a variance roughly proportional to $1/l^{\theta}(t)$. Since reliable estimation of variance parameter always requires a lot of data, we want to have a submodel for the variance so that only one parameter remains to be estimated for this submodel by the final model — analogous to simple linear regression. Thus, a good first choice for the observation model would be

$$ {}^{\text{obs}}G^{\theta}(t) \sim \text{normal}(G^{\theta}(t), \sigma^2/l^{\theta}(t)) \qquad \text{with a common } \sigma \text{ for all } \theta \in \Theta. $$

There may be other factors that affect the variance. For example, Siegel [4] suggests that the variance is also proportional to a power of the expected claim size, citing examples with a power of 1.5. Unfortunately, there are not enough studies on this yet to make a general statement, so for now it has to be examined on a case–by–case basis. Below, we apply our model to a set of similar inhouse products. For these, their observed variance is examined in the Appendix A. There we show that the best model for the observation variance would be

$$ \text{Var}({}^{\text{obs}}G^{\theta}(t)) = \sigma^2(1+bt)^2/l^{\theta}(t), $$

i.e. the standard deviation scales directly with the inflation term of the model for the expected age–normalized claim size, but is independent of the scaling factors! Note also that this model is invariant under currency changes. Unfortunately, the latter formula leads to identifiability problems for the model, since not only reasonable inflations b lead to a likely model, but also those with huge b, since then the variance is also huge. As a consequence, we have to fix the inflation term in the variance formula to some number ι in advance, so we write

$$ \text{Var}({}^{\text{obs}}G^{\theta}(t)) = \sigma^2(1+\iota t)^2/l^{\theta}(t). $$

We will start our investigation with $\iota = 0$ and later compute a sensitivity by taking into account the inflation found in the first run. One might as well start with the folklore inflation of $\iota = 3\%$ and then stick with it, since there is a large uncertainty in the variance estimate anyway, and the model should not be very sensitive to the inflation term.

After determining the variance structure, we move on to the correlation structure. Long–term or chronic illnesses of the insured will induce correlations in their claims in successive years, and hence temporal correlations between ${}^{\text{obs}}G^{\theta}(t)$. We expect the correlation to be stationary, i.e. $\text{Cor}({}^{\text{obs}}G^{\theta}(t), {}^{\text{obs}}G^{\theta}(t'))$ should depend only on the absolute time difference $|t-t'|$ and not on t or t' itself. In the Appendix B we show empirically that the following type of correlation matrices fit the data well:

$$ \begin{aligned} C(\rho,\lambda) &= (1-\lambda)\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots \\ \rho & 1 & \rho & \rho^2 & \ddots \\ \rho^2 & \rho & 1 & \rho & \ddots \\ \rho^3 & \rho^2 & \rho & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} + \lambda \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \\ &= (c_{tt'}) \qquad \text{with} \quad c_{tt'} = (1-\lambda)\cdot\rho^{|t-t'|} + \lambda\cdot 1, \quad \text{for} \quad \rho,\lambda \in [0,1[. \end{aligned} $$

This is a convex linear combination of a first–order autoregressive process correlation (also known as exponential correlation) and the fixed degenerate correlation of 1. Since a positive linear combination of a positive definite and a positive semidefinite matrix is positive definite, $C(\rho,\lambda)$ is in fact a correlation matrix. This convex linear combination also has a technical interpretation. We have a decaying correlation, which corresponds to the long–term illnesses or treatments that extend over the year–end period, and a

fixed permanent correlation, which corresponds to the chronic illnesses. Not surprisingly, we find that the correlation is relatively high for outpatient insurance, but relatively low for inpatient and dental insurance.

We will use the convention that omitting the dependence of a variable on t means that we consider the variable to be a vector for all t involved. Then we can summarize the **basis model** as

$$
\begin{aligned}
&\text{random observables} : {}^{\text{obs}}G^\theta \\
&\text{non random observables} : l^\theta \\
&\text{fixed values} : \iota \in \mathbb{R};\ \rho, \lambda \in [0,1[\\
&\\
&\text{to estimate: } a^\theta, b, \sigma \text{ such that with the definitions} \\
&\qquad G^\theta(t) := a^\theta(1+bt) \\
&\qquad \Sigma^\theta \quad := (\Sigma^\theta_{tt'}) \quad \text{with} \quad \Sigma^\theta_{tt'} := \sigma^2 \frac{(1+\iota t)\left((1-\lambda)\rho^{|t-t'|}+\lambda\right)(1+\iota t')}{\sqrt{l^\theta(t) l^\theta(t')}} \\
&{}^{\text{obs}}G^\theta \sim \text{normal}(G^\theta, \Sigma^\theta).
\end{aligned}
$$

Such a model can be fitted in the frequentist way with the function *gnls* of the **R** [18] package nlme [19] or in the Bayesian way with Stan [20], which we will do in the example below. Note that we need to know ρ, λ beforehand, they cannot be estimated — we would run into an identifiability problem. The reason is that we have only one time series of observations per product and also one free parameter a^θ per product, so the model cannot distinguish between a correlated random shift of the whole time series and a different value of the parameter a^θ.

Depending on the situation, it may be useful to develop the model further. So far we have assumed that all products have the same relative inflation b which is equivalent to them being similar. However, we might expect their relative inflations to be close but not equal because their benefits differ slightly. So we don't want to use one common relative inflation for all producs, but instead we want the product's inflation to be a mixture of such a common inflation and its own inflation. The more insured a product has, the greater the explanatory power of its own data and the more its own inflation should be weighted in this mixture. Such ideas are known in the actuarial context as credibility theory [21]. Nelder and Verrall [22] as well as Frees, Young and, Luo [23] showed that several important credibility models are equivalent to random effects modeling in the frequentist setting [24] respectively hierarchical modeling in the Bayesian setting [25]. There, each product is assumed to have its own relative inflation b^θ, but the inflations themselves are samples from a population distribution, normally distributed around a mean β, i.e. $b^\theta \sim \text{normal}(\beta, \sigma_b^2)$. This means that b^θ will be close to β unless the data provides sufficient evidence for a deviation. Thus in contrast to the base model in the **credibility / hierarchical model** we have

$$
\begin{aligned}
&\text{to estimate: } a^\theta, b^\theta, \beta, \sigma_b, \sigma \text{ such that with the definitions} \\
&\qquad G^\theta(t) := a^\theta(1+b^\theta t) \\
&\qquad \Sigma^\theta \quad := (\Sigma^\theta_{tt'}) \quad \text{with} \quad \Sigma^\theta_{tt'} := \sigma^2 \frac{(1+\iota t)\left((1-\lambda)\rho^{|t-t'|}+\lambda\right)(1+\iota t')}{\sqrt{l^\theta(t) l^\theta(t')}} \\
&b^\theta \quad \sim \text{normal}(\beta, \sigma_b^2) \\
&{}^{\text{obs}}G^\theta \sim \text{normal}(G^\theta, \Sigma^\theta).
\end{aligned}
$$

We can view each b^θ as the sum of the general inflation β and an idiosyncratic variation $b^\theta - \beta$ specific to this product, which is frequentist way of viewing this as an random effects model.

Technically, the frequentist way can be done using the *nlme* command of the R package nlme. It should be noted that fitting a hierarchical model can be computationally difficult [20, Section Reparameterization, 20, Section Gaussian Processes, 26, 27, Section 5.7, 28–30], especially when the variance of the hierarchical mean, here σ_b, becomes small, as we hope. The reason is that the full likelihood function gets some numerically unpleasant properties. However, these difficulties could be overcome in the Bayesian setting by using the techniques in the references, reparameterization, weakly informative prior, and reducing the step size.

Of course, we could take this one step further by modeling the product scaling factors a^θ hierarchically as well. This would apply credibility theory ideas to the claim size level of the products. However, we must be careful, because the underlying assumption of these methods is that the a^θ — and thus essentially the products themselves — are interchangeable or indistinguishable prior to looking at the claims experience. In other words, we should have no prior knowledge that the products should behave differently. In practice, we most likely know that some products have special characteristics or a long history of different claims experience, so this assumption may be violated. However, there are situation where this would be useful. For example, suppose you have designed a company health insurance plan and sold it to different companies. Each company should be self–funding, and we have no reason to assume in advance that the policyholders of one company will have higher claims than those of another. This is the ideal situation to use credibility methods and thus hierarchical modeling. This would prevent over– and under–pricing due to random fluctuations, as all prices are driven to a common mean until enough evidence is gathered to do otherwise.

4 Application

We will apply the above model to a set of 9 similar outpatient products for women of the DKV, a large German health insurance company. To avoid complications due to the COVID–19 pandemic, we choose the year 2019 as the test year and take the view of the year 2018, i.e. 2018 corresponds to $t = 0$. We will base our estimate on the five years before $T = \{-5, \dots, -1\}$.

Our model also requires the number of insured as input, particularly in the future. We will use the real values, in practice they will have to be estimated. In general, this should not be difficult. Mostly, the number of insured determines the future observation variance, which is not needed for the practical calculation of premiums. However, since we also assume that the observation errors are correlated, it also affects the speed with which the deviation of the observed normalized claim from its expected mean value will decrease in the future.

The products are ordered by their number of insured, and their exposure is shown in Figure 1. It differs by a factor up to nearly 10. In particular, the smallest product T9 with only 60 to 100 insured will turn out to be interesting.

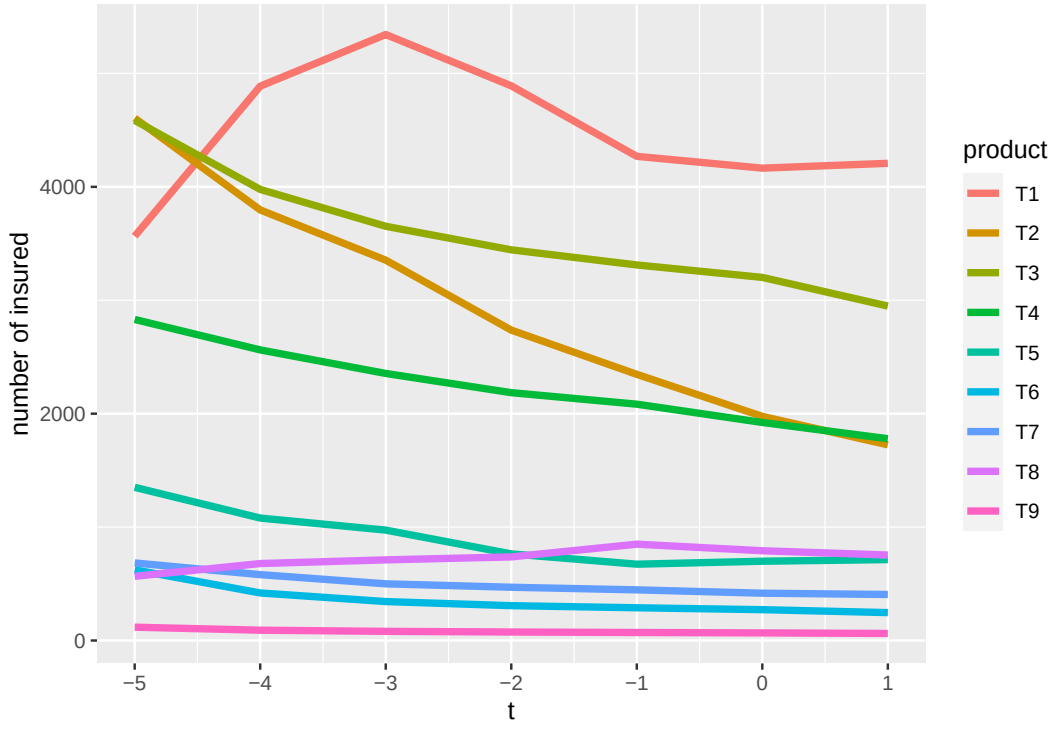


Figure 1: Exposure in outpatient products.

Before looking at the observed age–normalized claims, one must check for outliers that may violate the

normality assumption for the age–normalized claims too much. Here, there was no clear gap between the highest claims, and dropping those with the highest claims did not change the general behavior. However, there may be outliers, such as for men in these products. There is one claim of over 1.5 million and another insured has claims of over half a million for each year. Such cases have a serious impact on all values and should be dealt with accordingly.

In Figure 2 we plot their observed age–normalized claims, scaled by a common factor so that their median is approximately 1. The plot also includes their individual linear regression lines with an 80% confidence interval, based on observations for $t \in T$. It further contains the extrapolations two years into the future with their 80% prediction intervals, as well as the test points.

While this looks stable for several products, for some products the slope of their regression line is very uncertain. Also, overlaying all the regression lines in Figure 3 shows that the relative slope ranges from -1% to 5.8%, which is not plausible for similar products. Both of these observations are exactly the reason why we have introduced our model in the first place.

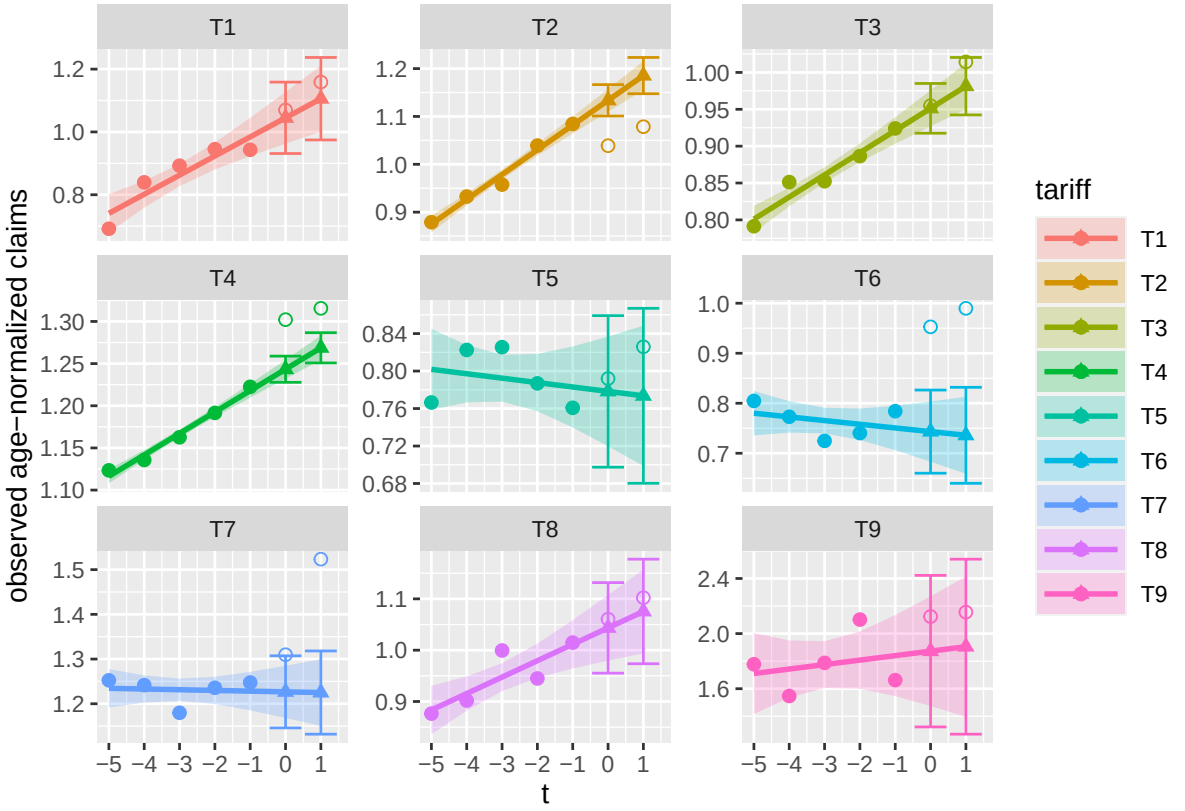


Figure 2: Separately fitted lines for $G^\theta(t)$ with 80% confidence intervals. The dots are the observed values used for fitting. The circles are the holdout observations. The triangles are predictions with 80% prediction intervals.

4.1 Non–Hierachical Models

To proceed with Bayesian modeling, we need to choose priors. We follow the advice of [27] and choose weakly informative priors rather than uninformative ones. This speeds up the computation and avoids unreasonable parameter values that we do not believe in anyway. Since the a^θ are positive, a lognormal distribution is the natural choice. Remembering that we have scaled all the claims so that the observed G have a median of about 1, we choose the parameters for the lognormal distribution so that the range from 0.5 to 3 is very likely. Folklore says that relative inflation b is about 3%. Since we consider the whole range from 0% to twice that to be very likely, we choose a normal distribution with mean 3% and standard deviation 3%. Finally, we need to choose a prior for the observation standard deviation factor σ. We first scale the fixed part of Σ with $\min_{\theta,t}\{l^\theta(t)\}$. Then σ has the same order of magnitude as a^θ and b. We choose a half–normal distribution for its square σ^2 as a prior. Its standard deviation is found

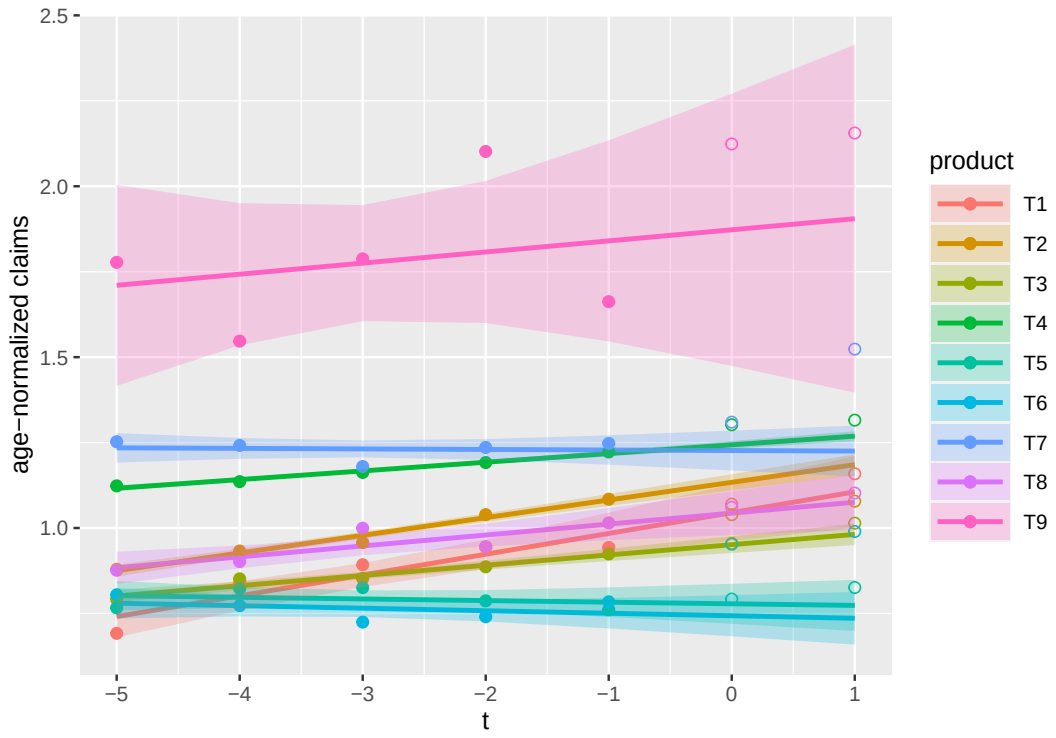


Figure 3: Separately fitted lines for $G^\theta(t)$ with 80% confidence intervals. The dots are the observed values used for fitting. The circles are the holdout observations.

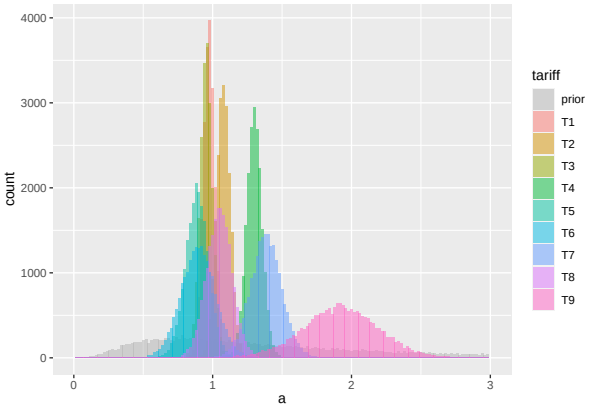
by the method described in [27]: simulate the model without data and choose parameters such that the implied distributions for the ${}^{\text{obs}}G^\theta(t)$ seem reasonable. In summary, we have as priors:

$$\begin{aligned} a^\theta &\sim \text{lognorm}(0.25, 0.75^2) \\ b &\sim \text{normal}(0.03, 0.03^2) \\ \sigma^2 &\sim \text{half–normal}(0, 0.5^2) \end{aligned}$$

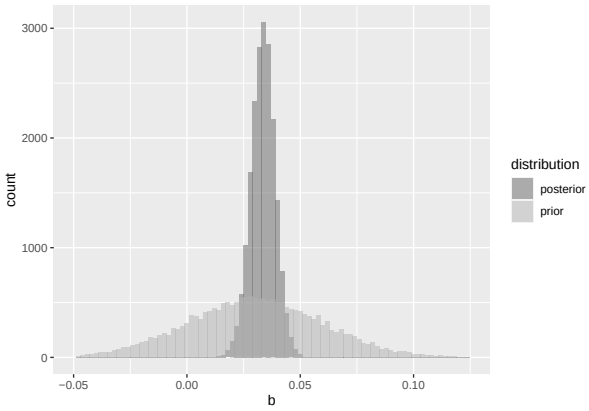
The Stan code for the model is supplied as an electronic supplement. For code development, the multi–normal distribution examples in the Stan manual [20, Section Gaussian Processes] and the Gaussian processes example by Lemonie [31] were helpful.

Figure 4d shows the fitted lines for the expected age–normalized claims together with the observed values, the fitting and test points. As desired, the lines all have the same relative slope and do not feel as random as in Figure 3. In Figure 4b and Figure 4c we see that the priors for b and σ^2 were indeed weakly informative, and that the data contain enough information to cluster the posterior distributions around a particular value. Figure 4a shows that this is less the case for a^θ. Because for each a^θ we can only use the data that the specific product θ provides, its distribution is less focused for the smaller products. This is especially the case for the smallest product T9. Unfortunately, its a^9 is also much higher than the others, and we expect its posterior to be pulled down by the prior, since the prior gives less weight to the higher values. Also the fitted line seems to lie too low with respect to the fitting points. Now we have to decide whether we consider the higher claims of the product T9 to be probable or not. Here we could confirm that product T9 does indeed have a long history of higher claims than the others. It is therefore undesirable for a^9 to be pulled down by the prior. In fact, the test points show that we underestimate the future claims.

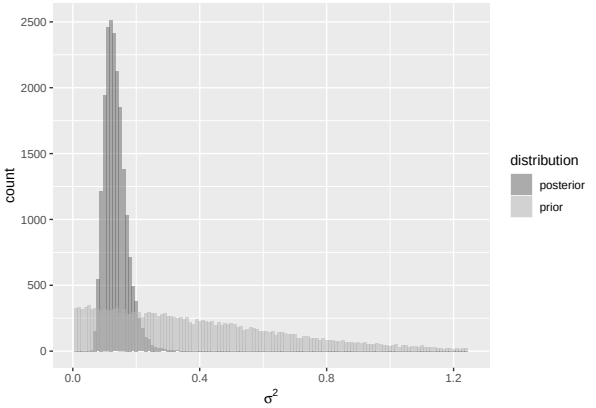
We can either exclude the product T9 from this set of products or choose a new prior. We do the latter and choose a uniform distribution in the range of 0.25 to 4 as a prior for a^θ, in summary:



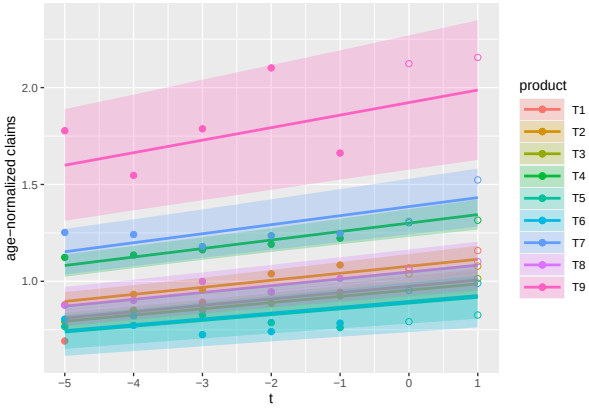
(a) Distribution of a^θ.



(b) Distribution of b.



(c) Distribution of σ^2.



(d) Fitted lines for $G^\theta(t)$ with 80% credibility intervals. The dots are the observed values used for fitting. The circles are the holdout observations.

Figure 4: Model: a lognormal b pooled.

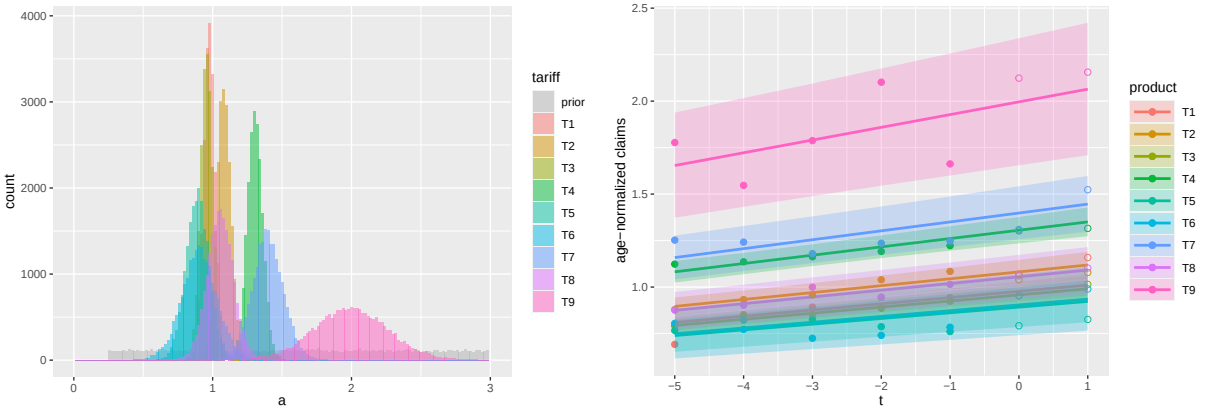
$$
\begin{aligned}
a^\theta &\sim \text{uniform}(0.25, 4) \\
b &\sim \text{normal}(0.03, 0.03^2) \\
\sigma^2 &\sim \text{half-normal}(0, 0.5^2)
\end{aligned}
$$

In Figure 5a we see the new posteriors for the a^θ. The one for product T9 has moved to the right, as intended. Indeed, the fitted line for product T9 now lies exactly between all the fitting points, see Figure 5b. Table 1 at the end of this section contains the posterior means of all the model variants we will discuss. There we see that among the other a^θ only those of the products with fewer insured are slightly affected by this model change. The prior/posterior plots for b and σ^2 are so similar to those of the previous model variant that we will not reproduce them here.

To judge the quality of our model, we show in Figure 6 the fitted lines in separate plots. We see that the regression lines lie in the "middle" of the 5 fitting points. However, unlike fitting the regression separately for products, we have forced the lines to have a similar relative slope, so it may well be that the first points are all on one side of the line and the remaining points are on the other. The figure also shows the effect of modeling correlations: the predictions for a product θ no longer necessarily lie on the regression line of $G^\theta(t)$. Starting from the last observation, the predictions slowly move closer to the regression line over time, because the observation error of the last year has less and less influence.

The lines look much more plausible for the products with the fewest insured, and many more of the test points stay in the 80% prediction intervals compared to the individual fits in Figure 3. Thus, the model is indeed an improvement. However, the test point for the product with the most insured T1 is clearly

outside the prediction interval. This is because its slope was much steeper than that of the other products in the individual fitting. The forced identical relative slope for all products leads to an underestimation in this case. So we want to relax this condition a bit and enforce only similar slopes, so that product T1 can have a steeper relative slope.



(a) Distributions of the a^θ.

(b) Fitted lines for $G^\theta(t)$ with 80% credibility intervals. The dots are the observed values used for fitting. The circles are the holdout observations.

Figure 5: Model: a uniform b pooled.

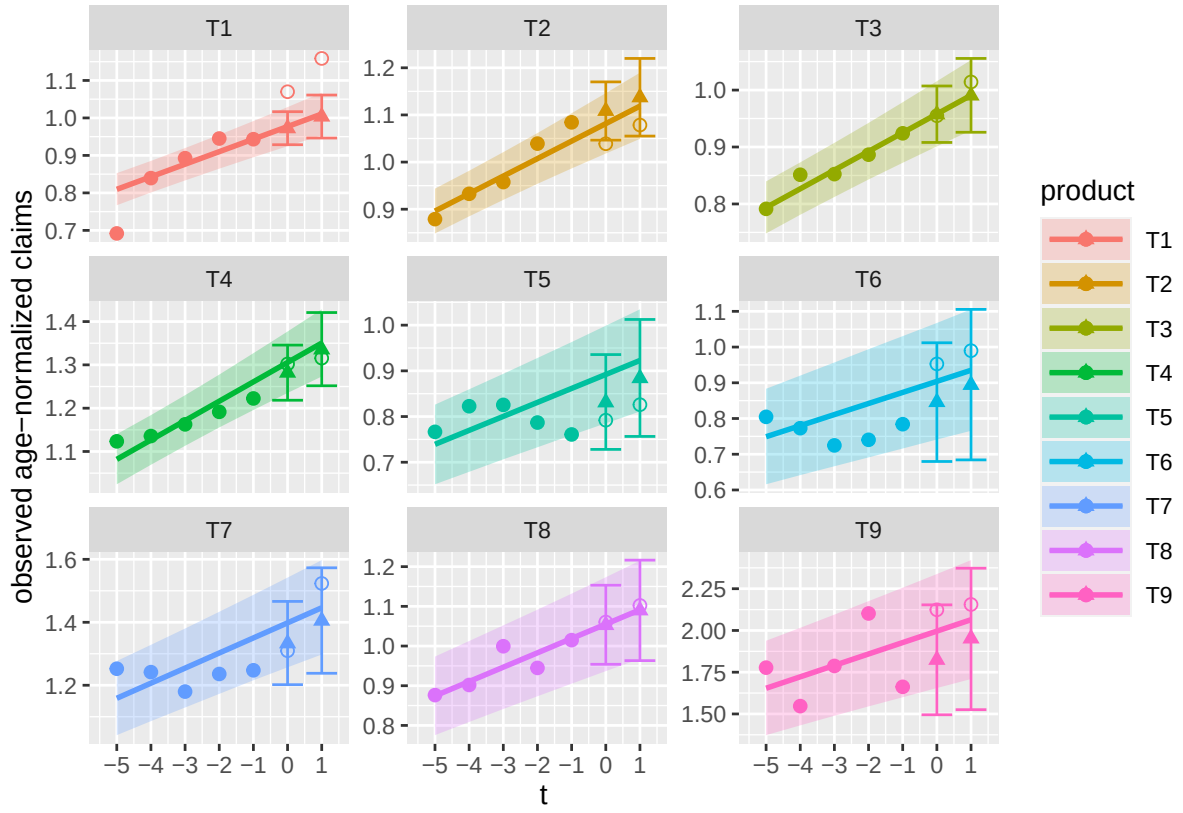


Figure 6: Model: a uniform b pooled: fitted lines for $G^\theta(t)$ with 80% credibility intervals. The dots are the observed values used for fitting. The circles are the holdout observations. The triangles are predictions with 80% credibility intervals.

4.2 Hierachical Models

So far, we have assumed that the relative slope b is the same for all products. This was due to our belief that the medical inflation should be the same for products with similar coverage. However, the products were sold under different circumstances and also have subtle differences in the coverage, so that we actually only expect them to have a very similar relative inflation, but not the same. We can model this using a hierarchical model. First, we model a parameter β as the "general mean" of all the individual relative slopes b^θ of the products. β is modeled exactly like b before. The b^θ are assumed to be close to β, technically they are drawn from a normal distribution with mean β and standard deviation σ_b, whose distribution we also have to choose. Since we want to make sure that the b^θ stay close together, we choose a prior for σ_b that is about half the time below 1%, we take the half–Cauchy distribution half-Cauchy$(0, 0.01)$. This is the first time we deviate from our general practice of choosing only weakly informative priors. To summarize:

$$\begin{aligned} a^\theta &\sim \text{uniform}(0.25, 4) \\ \beta &\sim \text{normal}(0.03, 0.03^2) \\ \sigma_b &\sim \text{half-Cauchy}(0, 0.01) \\ b^\theta &\sim \text{normal}(\beta, \sigma_b^2) \\ \sigma^2 &\sim \text{half-normal}(0, 0.5^2) \end{aligned}$$

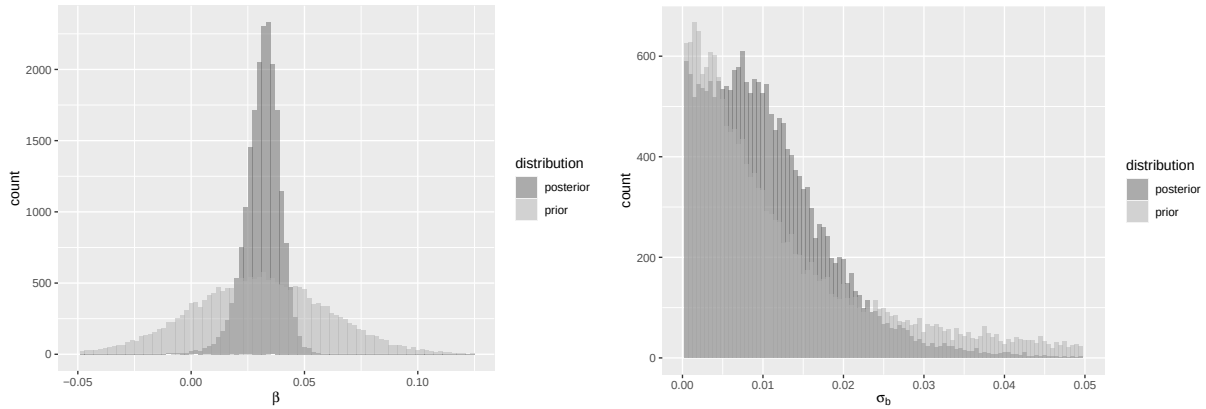
As mentioned at the end of Section 3, hierarchical models can cause computational difficulties, because the full likelihood function can have some numerically unpleasant properties. Here they could be overcome in the usual way by using a non–central parameterization, reducing the step size, and increasing the maximum tree depth. The Stan program for this model is again supplied as an electronic supplement. Figure 7a shows that β is very similar to the common b of the previous model. In Figure 7b we see that our prior for σ_b was informative, further the data suggest that larger differences between the b^θ are unlikely. Their histograms overlap to a large extent, so we show a close–up histogram in Figure 7c. We see that the b^θ have slightly separated from each other; this is most apparent for products whose individual relative slope is significantly different from the common one and which have many insured and therefore more reliable data, see Figure 3 and Table 1. In Figure 7d we see that the regression lines now have different relative slopes. We plot the regression lines individually in Figure 8. The test point for T1 is still outside the 80% prediction interval, but much less so.

As mentioned in the introduction, we could go one step further and model the a^θ hierarchically as well. But this would not fit our situation, since we would implicitly be assuming that we have no additional information that distinguishes the products, and we know that they will have different cost levels by their design. However, it may be useful in another context, so for demonstration purposes only, we also run a hierarchical fit on our data. We need to return to the lognormal prior for a^θ because a uniform distribution has no parameters that can be modeled hierarchically in a meaningful way. The following summary of priors includes a weakly informative choice for a^θ:

$$\begin{aligned} \alpha &\sim \text{lognorm}(0.25, 0.75^2) \\ \sigma_a &\sim \text{half-normal}(0, 0.75^2) \\ a^\theta &\sim \text{lognorm}(\log \alpha, \sigma_a^2) \\ \beta &\sim \text{normal}(0.03, 0.03^2) \\ \sigma_b &\sim \text{half-Cauchy}(0, 0.01) \\ b^\theta &\sim \text{normal}(\beta, \sigma_b^2) \\ \sigma^2 &\sim \text{half-normal}(0, 0.5^2) \end{aligned}$$

Some care had to be taken in choosing the priors in order to avoid too high values for a^θ, which would exceed the floating–point number limits of the computer. In particular, the typically chosen half–Cauchy distribution for a standard deviation does not work for σ_a because of its heavy tail, so the half–normal distribution was chosen instead, since we do not expect very high a^θ anyway.

As expected, if we compare the histograms of a^θ of this model variant, Figure 9a, with the histograms of models, in which the a^θ were modeled unpooled, see Figure 4a or Figure 5a, we see that the a^θ have



(a) Distribution of β.

(b) Distribution of σ_b.

(c) Histogram of the b^θ.

(d) Fitted lines for G^θ with 80% credibility intervals.

Figure 7: Model: a uniform b hierachical.

been pulled together. This implies that the same is true for the regression lines, compare Figure 9b with Figure 7d. In this new model variant, we see that a regression line can be completely above or below all observed values of that product. Of course, we could have chosen an even stronger prior for σ_a, giving more weight to the values near zero to bring them even closer together if desired.

4.3 Comparison and Sensitivities

Table 1 shows the means of the most important values for the models we have discussed so far. In addition, some sensitivities were computed for our favorite model "a uniform, b hierarchical". For "$\rho+, \lambda+$" the assumed correlation between the observed values in successive years was increased from $(\rho, \lambda) = (0.6, 0.25)$ to $(0.7, 0.35)$. For "$\iota+$" we assume a yearly increase in observation standard deviation of $\iota = 4\%$ up from 0%. The remaining sensitivities challenge the claim that our priors are only *weakly* informative. "σ^2 var+" widens the prior for the observation variance σ^2 from half-normal$(0, 0.5^2)$ to half-normal$(0, 0.75^2)$. "β mean +" shifts the prior for β to higher values from normal$(0.03, 0.03^2)$ to normal$(0.05, 0.03^2)$. "β, b^θ var+" widens the priors for β, b^θ from on $\beta \sim$ normal$(0.03, 0.03^2)$ and $\sigma_b \sim$ half-Cauchy$(0, 0.01)$ to $\beta \sim$ normal$(0.03, 0.05^2)$ and $\sigma_b \sim$ half-Cauchy$(0, 0.02)$.

By far the greatest impact results from the assumption that the relative inflation of all these products is the same or at least similar. Since the various model choices have already been discussed above, we will focus here on the sensitivities, in particular their effect on the predictions and thus on the pricing of the products.

The fixed parameters of the model are the annual increase ι in the observation variance, which causes the newer observations to be weighted less, and the correlation parameters ρ, λ. The sensitivities affect the predictions by less than 1%. In fact, the change remains below 0.5%, except for the products T1 and T6. The reason why these products are most affected is that their individual relative inflation deviates

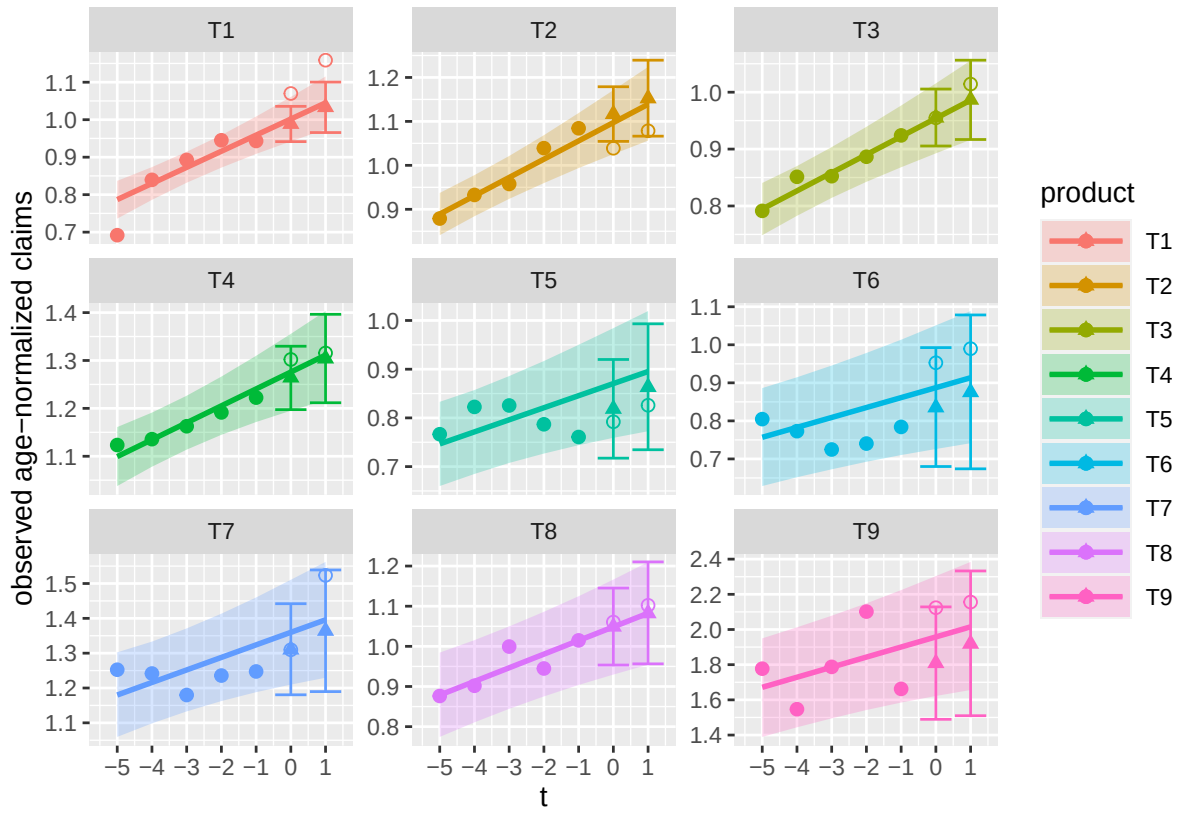
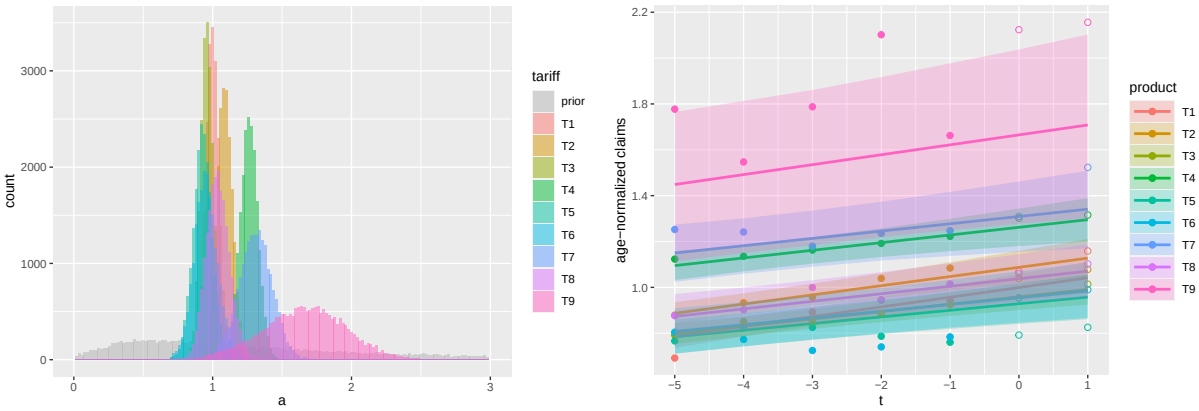


Figure 8: Model: a uniform b hierarchical: fitted lines for $G^\theta(t)$ with 80% credibility intervals. The dots are the observed values used for fitting. The circles are the holdout observations. The triangles are predictions with 80% credibility intervals.



(a) Distribution of the a^θ.

(b) Fitted lines for $G^\theta(t)$ with 80% credibility intervals.

Figure 9: Model: a hierarchical b hierarchical.

most from the common one, so their predictions are most affected by our model choice and therefore most affected by the choice of parameters. Of course, the minimal influence of these parameters on claim estimation is crucial for practical application, as their estimation is difficult and involves some uncertainty. Hence, it would not be sensible for the premium calculation of the products to heavily depend on them.

The sensitivities "σ^2 var+" and "β mean +" affect the predictions hardly at all, confirming that we have chosen the appropriate priors weakly informative. Finally, increasing the variance in the sensitivity "β, b^θ var+" changes the predictions by up to 1% as the σ_b prior contained stronger a priori assumptions.

Again, all these effects are dwarfed by our assumption that the products have similar relative inflation,

Table 1: Sensitivity of the parameter means

model / sensitivity	T1	T2	T3	T4	T5	T6	T7	T8	T9
a^θ									
individual	1.045	1.134	0.951	1.243	0.778	0.743	1.227	1.043	1.873
a lognorm b pooled	0.974	1.078	0.955	1.301	0.888	0.896	1.386	1.048	1.923
a uniform b pooled	0.977	1.082	0.958	1.306	0.892	0.904	1.398	1.055	1.996
a uniform b hierar.	1.002	1.097	0.954	1.276	0.870	0.887	1.360	1.048	1.958
— $\rho+, \lambda+$	1.011	1.092	0.954	1.281	0.877	0.911	1.376	1.050	1.985
— $\iota+$	1.008	1.097	0.956	1.279	0.883	0.904	1.376	1.051	1.996
— σ^2 var+	1.002	1.099	0.954	1.275	0.871	0.888	1.359	1.049	1.957
— β mean+	1.002	1.097	0.954	1.276	0.870	0.887	1.360	1.048	1.958
— β, b^θ var+	1.008	1.103	0.953	1.270	0.863	0.879	1.347	1.048	1.944
b^θ									
individual	0.058	0.046	0.032	0.020	-0.006	-0.010	-0.001	0.031	0.017
a lognorm b pooled	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
a uniform b pooled	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
a uniform b hierar.	0.043	0.038	0.033	0.027	0.028	0.029	0.026	0.032	0.029
— $\rho+, \lambda+$	0.041	0.037	0.033	0.029	0.029	0.030	0.027	0.033	0.030
— $\iota+$	0.046	0.038	0.034	0.027	0.028	0.028	0.025	0.033	0.029
— σ^2 var+	0.043	0.038	0.033	0.027	0.028	0.029	0.026	0.032	0.029
— β mean+	0.043	0.038	0.033	0.027	0.028	0.029	0.026	0.032	0.029
— β, b^θ var+	0.045	0.039	0.033	0.026	0.026	0.026	0.023	0.031	0.026
${}^{\text{obs}}G^\theta(1)$									
observed	1.159	1.079	1.014	1.316	0.826	0.990	1.523	1.102	2.156
individual	1.106	1.185	0.981	1.269	0.774	0.736	1.225	1.075	1.905
a lognorm b pooled	1.001	1.135	0.988	1.332	0.882	0.889	1.397	1.085	1.925
a uniform b pooled	1.004	1.137	0.990	1.336	0.884	0.894	1.405	1.090	1.952
a uniform b hierar.	1.033	1.152	0.987	1.304	0.863	0.875	1.364	1.082	1.919
— $\rho+, \lambda+$	1.030	1.154	0.988	1.307	0.857	0.875	1.364	1.084	1.897
— $\iota+$	1.041	1.157	0.989	1.304	0.867	0.879	1.365	1.086	1.925
— σ^2 var+	1.033	1.154	0.987	1.303	0.863	0.876	1.364	1.083	1.914
— β mean+	1.033	1.152	0.987	1.304	0.863	0.875	1.364	1.082	1.919
— β, b^θ var+	1.041	1.158	0.986	1.298	0.856	0.866	1.353	1.080	1.899

so they are less important. Also, the changes in the predictions in the sensitivities are largest for the products where the individual inflation differs most from the common inflation.

5 Conclusion

We have proposed a model for the claims development of a set of similar products. The key ingredient was the assumption that they all have the same or at least similar relative inflation. This model has been successfully applied to an example. For most products with fewer insured and therefore more volatile claims experience, this was the only way to make a meaningful forecast. The model relies on a detailed submodel for the observation error of the expected age–normalized claims. Further, it was shown that the claims predictions in the example were not very sensitive to the parameters of this submodel, so in practice these parameters could probably be fixed for large groups of products.

The model was developed in the Bayesian framework, which allowed us to incorporate our beliefs. We chose only weakly informative priors, except in a credibility variant of the model in which we modeled relative inflation hierarchically. There we wanted to bring the relative inflations closer together, but the data only confirmed our a priori expectations and the posterior was essentially the same as the prior. For the products with the most insured the difference in the relative inflation between these models was noticible. In practice, one must decide whether the increased level of detail in the hierarchical model is worth the added complexity and computational cost. It will be most useful when there are several products with a medium number of insured, because only then there can be sufficient evidence for different

relative inflation.

6 References

1. GKV-WSG (2007) Gesetz zur Stärkung des Wettbewerbs in der Gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz — GKV–WSG) (G–SIG: 16019322)
2. European Court of Justice (2011) Case C-236/09, sex discrimination in insurance premiums
3. Bierth K, Dietrich K, Jahnke D, et al (2023) Fachgrundsatz der Deutschen Aktuarvereinigung e.V. — Kalkulation und Bestandsgröße in der privaten Krankenversicherung — Hinweis
4. Siegel G (1997) Varianzschätzungen von Kopfschäden. Blätter der DGVFM 23:147–172
5. Milbrodt H, Röhrs V (2016) Aktuarielle Methoden der deutschen Privaten Krankenversicherung. Verlag Versicherungswirtschaft
6. Gottschalk T, Lax C (2019) Internal documents of the DKV (Deutsche Krankenversicherung)
7. Käsgen H (2023) Stützverfahren für die Berechnung des Grundkopfschadens in der Privaten Krankenversicherung. Bachelor's thesis, Rheinisch–Westfälisch Technische Hochschule Aachen
8. VVG (2024) Versicherungsvertragsgesetz vom 23. November 2007 (BGBl. I S. 2631), das zuletzt durch Artikel 4 des Gesetzes vom 11. April 2024 (BGBl. 2024 I Nr. 119) geändert worden ist.
9. KVAV (2018) Krankenversicherungsaufsichtsverordnung vom 18. April 2016 (BGBl. I S. 780), die zuletzt durch Artikel 6 Absatz 9 des Gesetzes vom 19. Dezember 2018 (BGBl. I S. 2672) geändert worden ist.
10. Becker T (2017) Mathematik der privaten Krankenversicherung. Springer Spektrum Wiesbaden
11. Gertrud Jäger (1958) Die versicherungstechnischen Grundlagen der deutschen privaten Krankheitskostenversicherung. Schriftenreihe des Instituts für Versicherungswissenschaft an der Universität Köln, Neue Folge, Heft 15. Duncker & Humblot Verlag
12. Rusam F (1936--39) Mathematik der Krankenkostenversicherung. In: Lencer R, Riebesell P (eds) Deutsche Versicherungswirtschaft ein Unterrichts- u. Nachschlagewerk 4 Versicherungsmathematik. Berlin : Verlag der Betriebswirt Franke & Co. K.G., pp 225–252
13. Buchner F, Wasem J (2000) Versteilerung der alters- und geschlechtsspezifischen Ausgabenprofile von Krankenversicherern. Zeitschr f d ges Versicherungsw 89:357–392
14. Christiansen M, Denuit M, Lucas N, Schmidt J-P (2018) Projection models for health expenses. Annals of Actuarial Science 12:185–203
15. Piontkowski J (2020) Forecasting health expenses using a functional data model. Annals of Actuarial Science 14:72–82
16. BaFin (2018) Wahrscheinlichkeitstafeln PKV 2017
17. Behne J (1993) Über den wahren und den rechnungsmäßigen Kopfschaden in der Krankheitskosten-Versicherung. Blätter der DGVFM 21:133–140
18. R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
19. Pinheiro J, Bates D, R Core Team (2023) nlme: Linear and Nonlinear Mixed Effects Models
20. Team SD (2024) Stan Modeling Language Users Guide and Reference Manual, 2.33
21. Bühlmann H, Gisler A (2005) A Course in Credibility Theory and its Applications. Springer–Verlag Berlin Heidelberg
22. Nelder JA, Verrall RJ (1997) Credibility Theory and Generalized Linear Models. ASTIN Bulletin 27:71–82
23. Frees EW, Young VR, Luo Y (1999) A longitudinal data analysis interpretation of credibility models. Insurance: Mathematics and Economics 24:229–247
24. Zuur AF, Ieno EN, Walker N, et al (2009) Mixed Effects Models and Extensions in Ecology with R. Springer–Verlag New York

25. Gelman A, Carlin JB, Stern HS, et al (2013) Bayesian Data Analysis (3rd ed.). Chapman; Hall/CRC
26. Chung Y, Rabe–Hesketh S, Dorie V, et al (2013) A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models. Psychometrika 78:685–709
27. Gelman A, Vehtari A, Simpson D, et al (2020) Bayesian Workflow
28. Betancourt M, Girolami M (2015) Hamiltonian Monte Carlo for Hierarchical Models. In: Upadhyay SK, Singh U, Dey DK, Loganathan A (eds) Current Trends in Bayesian Methodology with Applications. Chapman and Hall/CRC
29. Betancourt M (2017) Diagnosing Biased Inference with Divergences. https://betanalpha.github.io/assets/case_studies/divergences_and_bias.html. Accessed 9 Jan 2024
30. Betancourt M (2020) Hierarchical Modeling. https://betanalpha.github.io/assets/case_studies/hierarchical_modeling.html. Accessed 9 Jan 2024
31. Lemoine N (2017) Even Faster Gaussian Processes in STAN. https://natelemoine.com/even-faster-gaussian-processes-in-stan/. Accessed 20 Mar 2024
32. Binder M (1998) Das Schwankungs- und Schätzrisiko. In: Transactions of the 26th International Congress of Actuaries, Birmingham, UK. Institute of Actuaries Britain, pp 55–72

A Empirical Observation Variance

The observation error of the age–normalized claims is an essential part of any sophisticated claims development model. However, estimating its variance is sensitive to outliers, so it is desirable to have a submodel for it that combines observations from multiple years or even different products to reduce the effect of outliers. We expect the claims development model itself to be not very sensitive to specific details of this submodel. However, this must be justified by a sensitivity analysis, as we have done in Section 4.3. Our goal here is to develop a submodel that defines the relative relationship of all occurring observation variances to each other. The determination of a final common factor is left to the main model.

Recall that we assume that the observed age–normalized claims ${}^{\text{obs}}G^{\theta}(t)$ are normally distributed with mean $G^{\theta}(t)$. Since ${}^{\text{obs}}G^{\theta}(t)$ is itself a weighted average of $l^{\theta}(t)$ individual claims, it is natural to assume that $\text{Var}({}^{\text{obs}}G^{\theta}(t))$ is inversely proportional to $l^{\theta}(t)$, holding everything else fixed. Unfortunately, we already know that the claims of the individuals are not identically distributed, mainly because of their obvious age dependence. However, if the age distribution is the same — or at least very similar — over time and over the different products, we can consider $l^{\theta}(t)$ as an overall scaling factor, and the relationship will still hold. We will return to age dependence at the end of this section.

When modeling an average loss in insurance, it is often assumed that there is a power law between the variance and the mean. In a context similar to ours, this is already discussed by Siegel in [4, Section 7.3.1], who refers to an older text by Binder that was later published in [32]. In our notation, this would suggest

$$\text{Var}({}^{\text{obs}}G^{\theta}(t)) = A^{\theta}\frac{1}{l^{\theta}(t)}(G^{\theta}(t))^{\alpha} \quad \text{for suitable } A^{\theta}, \alpha \geq 0.$$

Siegel notes that this probably only makes sense for $1 \leq \alpha \leq 2$. He also mentions unpublished empirical studies with α around 1.5. He also points out that $\alpha = 2$ is the most theoretically pleasing variant, because only then the proportionality constants A^{θ} are currency independent. We also note that we get this law with $\alpha = 2$ if the changes in $G^{\theta}(t)$ are induced by a simple year only dependent scaling of each individual claim.

To derive such a law, we need an estimator for the variance. Siegel uses the standard estimator. Because of the age dependence of the claims he must apply it separately to each age group and then combine the results to obtain a variance estimate of the age–normalized claim. This method has two minor drawbacks. First, there will be age groups that have only few insured and thus their variance estimate will be unreliable; fortunately, these are given less weight in the estimate of the variance of the age–normalized claim. Second, it is unclear how to deal with lapses and product changes during the year rather than at the end of the year. Simply scaling the claims to a full year will easily produce very large claims and thus overestimate the variance in this age group. Ignoring them when estimating the variance is also not a

good option either, because the large near–death claims are among these claims that should be taken into account.

Now that we have more computing power, we can use naive bootstrapping, i.e. we consider all of the insured as typical representatives of the product. Then we resample the insured of each product 5000 times with replacement, keeping the total number of insured fixed, i.e. each sample may contain several copies of an insured or none at all. For each sample set of insured, we take their respective claims and compute the age–normalized claim as usual. This gives us 5000 realizations of the age–normalized claim, which we can use to estimate its observation variance.

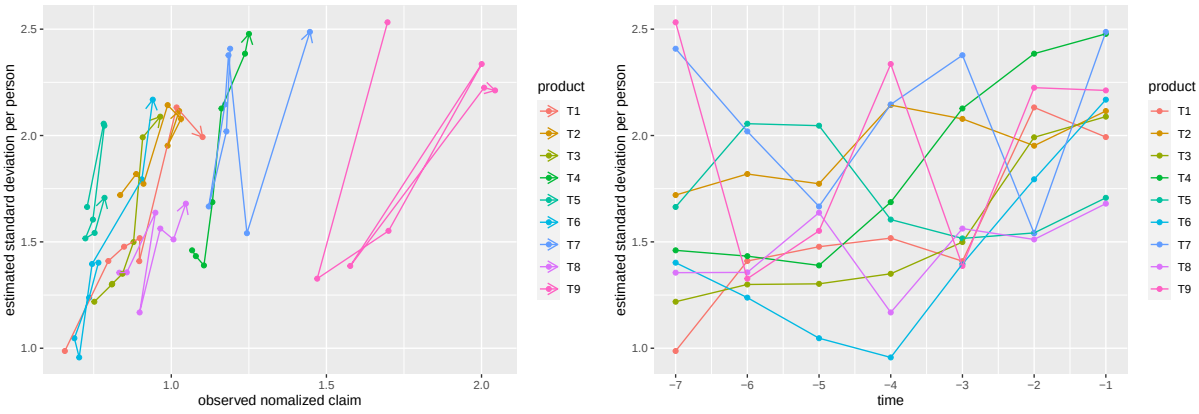
In Figure 10a we plot the "estimated standard deviation per person" $\sqrt{l^\theta(t)\hat{\mathrm{Var}}({}^{\mathrm{obs}}G^\theta(t))}$ against the ${}^{\mathrm{obs}}G^\theta(t)$ connecting successive observations of each product. The most striking feature is the apparent randomness, which confirms that the variance estimator is very volatile. There may be some increase in the standard deviation per person as the observed age–normalized claims increase, but this impression is mostly due to the points in the lower left corner, and these products have much higher variance in other years, so we can consider these points as random points as well.

However, we can discover a feature of the data if we follow the time evolution of each product separately. Here we note an increase in the estimated standard deviation per person over time, see Figure 10b. The relative increase, ι, is about 4% to 6% per year, about the size of the general relative claim inflation found in the main model. If we attribute everything else to randomness, we get the model

$$\sqrt{l^\theta(t)\mathrm{Var}({}^{\mathrm{obs}}G^\theta(t))} = (1+\iota t)\sigma \quad \Longleftrightarrow \quad \mathrm{Var}({}^{\mathrm{obs}}G^\theta(t)) = \frac{(1+\iota t)^2}{l^\theta(t)}\sigma^2 \quad \text{for a suitable } \sigma$$

where ι is approximately at the level of the relative inflation.

Note that for ι equal to the relative claim inflation, our main model implies that $G^\theta(t)$ is proportional to $1+\iota t$, so this formula is equivalent to the power law discussed at the beginning with $\alpha = 2$, i.e. the theoretically pleasing version with dimensionless proportionality constants A^θ. Somewhat surprisingly, we also find that the variance formula depends only through the number of insured, $l^\theta(t)$, on the particular product θ. Of course, we must remember that we are looking at a set of *similar* products; for non–similar products, the result will surely be different.



(a) the age–normalized claims connecting the consecutive observations of each product.

(b) time.

Figure 10: Standard deviation per person against ...

Finally, we briefly discuss how the standard deviation per person depends on the age structure. To do this, we repeat the bootstrapping process above, but this time we compute the age–normalized claim separately by age. Figure 11 shows the result for the four products with the most insured. We can see a bit of a frown there. In particular, all the higher values occur in the middle ages. So the age structure will be of some importance and some care should be taken when mixing products with very different age structures. The effect could be of the same order of magnitude as the effect of claims inflation over some years. Note, however, that in our example the exposures in the products vary by a factor of up to

100, resulting in an exposure effect of up to a factor of 10 between the different products, which clearly dominates the age structure effect.

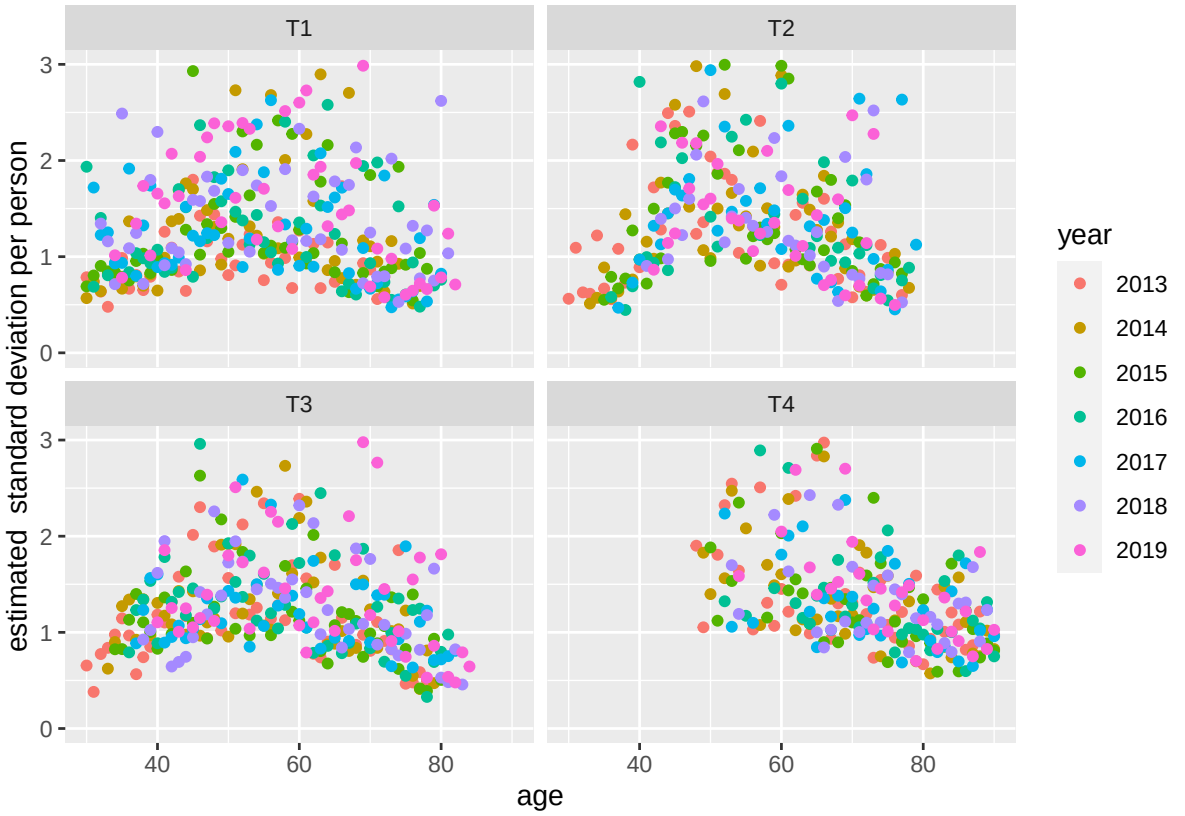


Figure 11: Estimated standard deviation of the age–normalized claims per person against age.

B Empirical Observation Correlation

Correlations between an insured's claims in successive years occur because of chronic or long–term illnesses or simply because the treatment period extends beyond the end of the year. Because this topic has been neglected in the actuarial literature on German health insurance, we will cover all product variants (outpatient, inpatient, and dental) to provide a reference. We have chosen products that have a large number of insured to provide stable estimates and that have a long history to provide information on long–term dependency.

Of course, the correlations will depend on the characteristics of the product. The products shown here have been on the market for a long time, so the average age of the insured is higher than in the newer products, hence we expect a higher correlation due to a higher prevalence of chronic diseases. This effect is offset by the fact that many insured changed to newer products, which reduces the correlation in both products. However, we believe that the values presented here are a useful starting point for examining other products. For the women in the outpatient insurance, in addition to this product here, we also examined the products in the main section and obtained similar values, confirming our belief. A simulation also showed that a medium deductible in a product has little effect on the correlations.

Unfortunately, outliers significantly affect the correlations. We removed three claims over one million from the inpatient products because there seemed to be a natural gap in claim sizes at around one million. For the outpatient products, there was no obvious choice. We decided to remove three men and two women who had claims over a quarter of a million for several years. This reduced the correlations for women by about 10%. Thus, when applying the results here in practice, one must consider how the particular insurance company deals with extremely high claims. Figure 12 shows the exposure in the products. They are high enough to expect reliable estimates.

Naturally, we expect that the correlation between two observed age–normalized claims, $\mathrm{Cor}({}^{\mathrm{obs}}G^{\theta}(t), {}^{\mathrm{obs}}G^{\theta}(t'))$,

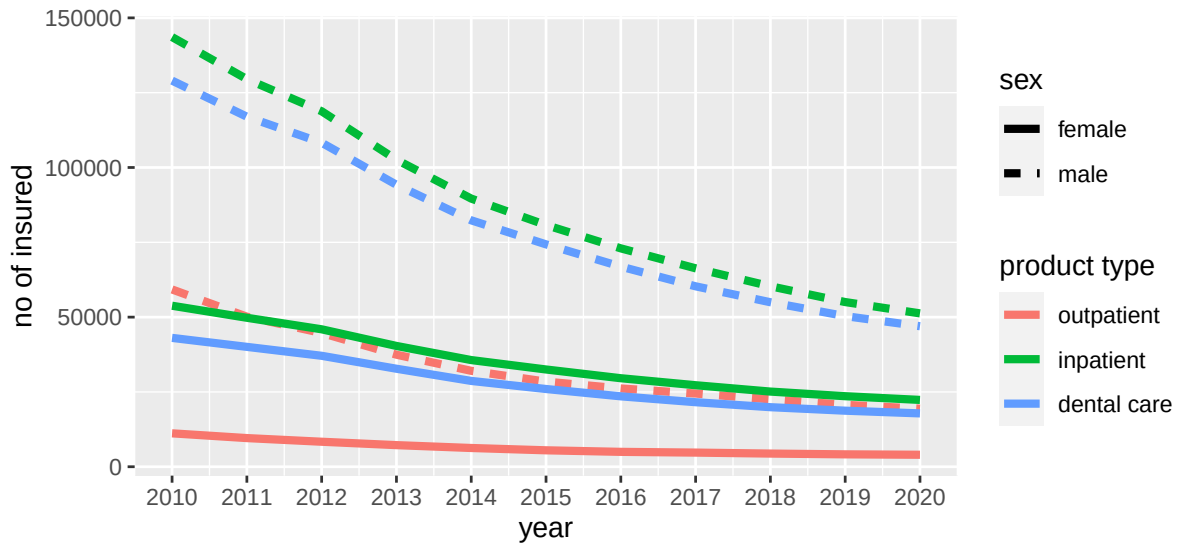


Figure 12: Exposure in the products.

Table 2: Parameters for correlation submodel

product type	sex	ρ	λ
outpatient	female	0.59	0.26
	male	0.62	0.26
inpatient	female	0.17	0.06
	male	0.17	0.05
dental care	female	0.18	0.07
	male	0.15	0.07

depends only on the time difference $\Delta t = |t' - t|$ and not on the particular t and t' themselves. We compute the correlation by bootstrapping like we did for the variance in the section before. Figure 13 shows the results as a function of t and Δt. We have also computed an average over t by first standardizing the bootstrapped age–normalized claims per year and then pooling all pairs with appropriate time difference together. Inpatient and dental care have a low correlation of less than 25% in neighboring years and it drops fast for larger time differences. In contrast the outpatient product shows a high correlation of about 70% in neighboring years which decays more slowly in time. Also note that we get similar results for males and females. As mentioned above the insured in these products are older, so in particular nearly all females are above childbearing age, thus the main reason for different correlations is absent here. Finally, we want to describe the correlation by a simple formula. Figure 13 suggest that on the one hand there is a fast, maybe exponential, decay and on the other hand that there is still some correlation after a long time. This suggest the ansatz

$$\mathrm{Cor}({}^{\mathrm{obs}}G^{\theta}(t), {}^{\mathrm{obs}}G^{\theta}(t')) = (1-\lambda)\cdot\rho^{|t'-t|} + \lambda\cdot 1.$$

This is a positive linear combination of the correlation structure of a first–order autoregressive process and a fixed correlation. Since a positive linear combination of a positive definite and a positive semidefinite matrix is positive definite, and since it's obviously equal to 1 for $t = t'$, this is indeed a correlation structure. Figure 13 contains the curves fitted to the bootstrapped correlations for all $(t, \Delta t)$. Table 2 contains the parameters for further reference. We get very good fits for outpatient insurance — where it matters most — and dental care. For inpatient insurance, a more complex function may be needed, but since the values, and thus the fitting errors, are small, it is unlikely to be important in practical applications.

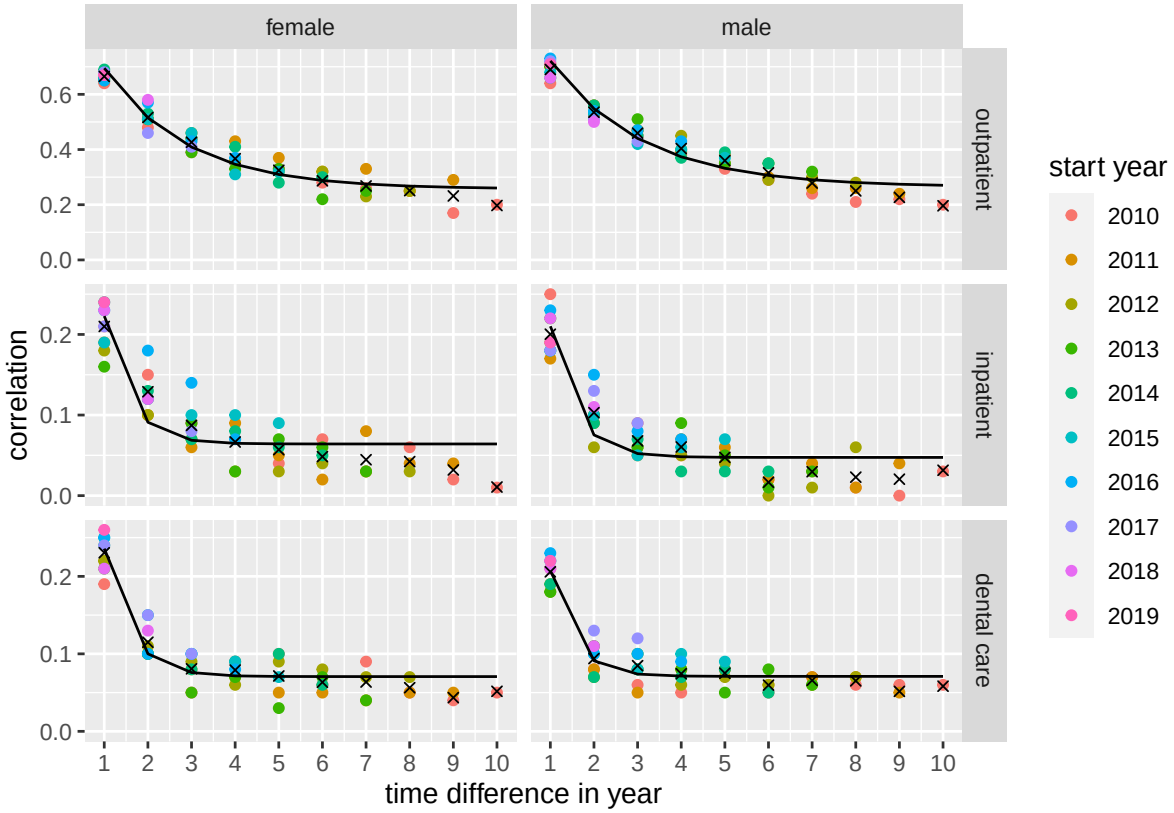


Figure 13: Correlations between years. Crosses are the averages. The lines are the fitted ansatz.