Pricing small tariffs in German health insurance

Jens Piontkowski[1,*]

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

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

Abstract

Traditionally, in German health insurance, a small tariff with volatile claims is priced using a similar large tariff as a reference. However, legislative changes and market forces have led to a fragmentation of tariffs. As a result, such a large similar tariff is often no longer available. Here we propose a statistical model that combines the data of several small tariffs to derive a common relative claim inflation as well as the expected claims of these tariffs in the future, thus enabling stable pricing for these tariffs.

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 tariffs in recent years. As a consequence the new tariffs have fewer insured than before and therefore their claim experience is more volatile. In particular, it is difficult to estimate the yearly increase of claims as it is often in the order of the statistical fluctuations. Previously, the yearly increase was often taken from a similar large older tariff (a technic known in German as "Stütztarif"), but many of the insured of these older tariffs changed to newer tariffs and these older tariffs have become less stable themselves. Thus the task is now to determine the yearly claim increase not from one tariff, but from a collection of similar tariffs.

Similar to the treatment of tariff classes by Milbrodt and Röhrs [3, Chapter 5], Gottschalk and Lax [4] used an ad hoc estimator to scale similar tariffs to a common level. Then the claims can be combined and a simple linear regression yields the desired annual claim increase, which is then scaled down to the individual tariffs. This solved the problem for the requirements of premium calculation in practice. In the thesis of Käsgen [5] 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, linear regression, 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 tariff 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 smaller tariffs.

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 parameter 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

We recall the essential part of German health insurance premium calculation, more precisely how to adjust the premiums of a large tariff which has been sold for several years. For the details that have to be considered in practice see Milbrodt and Röhrs [3, Chapter 5] or Becker [6, Section 3.2]. For notational convenience the current year is denoted by $t = 0$. We assume that we can use the experience of the former years $T = \{t_{\min}, \ldots, -1\}$ for the estimation of the claim size in the next year, $t = 1$, which is our goal. In practice $t_{\min}$ is often between -6 and -3, we will use -5 in the examples.

In German health insurance the insured are distinguished only by age and sex. The newer tariffs have a unisex premium, but the claim estimation is first done separately by sex, and only later are the sexes combined. Thus for our purpose we may assume the sex is fixed and only the age dependence remains. The age groups of children and teenagers are treated separately from the adults in a simplified way. Thus we focus on the adults and consider only a subset X of all possible ages. For the insured of age $x \in X$ in tariff θ we seek to estimate the mean claim per capita (in German “Kopfschaden”), $K_x^\theta(t)$, for the next year $t = 1$. The number of insured per age and year may be small and thus the observed average claims, ${}^{\text{obs}}K_x^\theta(t)$, may be heavily influenced by random fluctuations. To reduce the number of free parameters, Rusam introduced in the 1930s the assumption that the expected claims per year and age can be split 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 $G^\theta(t)$ an age normalized claim (in German “Grundkopfschaden”). In practice, this assumption is a good approximation for short time periods, in the long term age profiles change, see [7–9]. To make the parameters identifiable one needs an additional assumption; for example setting $k_{x_0}^\theta = 1$ for a reference age x_0. The general references mentioned at the beginning explain how to derive an age profile. However, often the example ones published yearly be the German government organisation BaFin are used, e.g. [10]. We will assume that the age profiles are given.

Denote by $l_x^\theta(t)$ the number of insured of age x in year t in tariff θ then we expect the following total claim per age:

$$S_x^\theta(t) := l_x^\theta(t) K_x^\theta(t) = l_x^\theta(t) k_x^\theta G^\theta(t),$$

and for the whole tariff

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

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

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

With these formulas our task to estimate the claim size of the next year comes down to estimate $G^\theta(1)$. In German health insurance, there is a tradition of assuming a linear development of the normalized claims:

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

Following Behne [11, Section 2], Milbrodt and Röhrs [3, 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, here it is not used in practice, the reasons for this may include, that this is only a short term projection and therefore the results do not differ much, that linear regression is the simpler method and has been in use for many years, or that it is even suggested for a similar purpose by the governmental ordinance KVAV [12, Anlage 2B].

${}^{\mathrm{obs}}G^\theta(t)$ 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 large tariff. This leads to a simple linear regression model, i.e. we have

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

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

3 The Model for a Set of Similar Tariffs

Now assume that instead of one large tariff θ with many insured, we have a set Θ of medium sized tariffs. In order to proceed these must be similar in some way. Inspired by the definition of a tariff class in [3, Section 5.8] we make the following

Definition 3.1 (similar tariffs)**.** A set of tariffs Θ with given age profiles is called *similar* tariffs, if and only if their 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 normalized claim* is

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

The definition of a tariff 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 tariffs 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 tariffs, so that proportionality also holds with the same factor for each claim size per age.

As mentioned above Gottschalk and Lax [4] introduced a natural ad hoc estimator for the proportionality factors. With these factors, all the claim experience can be scaled to the fictitious base tariff. Then one can combine all the claims and proceed as before for a large tariff. After extrapolating the fictitious base tariff one needs to undo the scaling to get back to the individual tariffs.

Here, we will combine all of these steps into one model that estimates everything in a single step. Assuming a linear trend for the fictive base 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 tariffs. We resolve this by requiring $1 = G(0) = a$. In summary, we assume that the expected normalized claims follow

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

When extrapolating a large tariff it is often assumed that the observation error is homoscedastic, because the number of insured in the tariff is nearly constant over time. Here we need to be more careful, because in general tariffs 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 [13] suggests that the variance is also proportional to a power of the average 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 tariffs. 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 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

random observables : ${}^{\text{obs}}G^\theta$
non random observables : l^θ
fixed values : $\iota \in \mathbb{R};\ \rho, \lambda \in [0, 1[$

to estimate: a^θ, b, σ such that with the definitions

$$
\begin{aligned}
G^\theta(t) &:= a^\theta(1 + bt) \\
\Sigma^\theta \quad &:= (\Sigma^\theta_{tt'}) \quad \text{with} \quad \Sigma^\theta_{tt'} := \sigma^2 \frac{(1+\iota t)\left((1-\lambda)\rho^{|t-t'|}+\lambda\right)(1+\iota t')}{\sqrt{l^\theta(t) l^\theta(t')}}
\end{aligned}
$$

$${}^{\text{obs}}G^\theta \sim \text{normal}(G^\theta, \Sigma^\theta).$$

Such a model can be fitted in the frequentist way with the function *gnls* of the **R** [14] package nlme [15] or in the Bayesian way with Stan [16], 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 tariff and also one free parameter a^θ per tariff, 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 tariffs 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. So we don't want to use one common relative inflation for all tariffs, but instead we want the tariff's inflation to be a mixture of such a common inflation and its own inflation. The later the more, the more evidence we have for it, i.e. the larger the tariff. Such ideas are known in the actuarial context as credibility theory [17]. Nelder and Verrall [18] as well as Frees, Young and, Luo [19] showed that several important credibility models are equivalent to random effects modeling in the frequentist setting [20] respectively hierarchical modeling in the Bayesian setting [21]. There, each tariff 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. We will show this in the example below. 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 [16, Section Reparameterization, 16, Section Gaussian Processes, 22, 23, Section 5.7, 24–26], 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 tariff scaling factors a^θ hierarchically as well. This would apply credibility theory ideas to the claim size level of the tariffs. However, we must be careful, because the underlying assumption of these methods is that the a^θ — and thus essentially the tariffs themselves — are interchangeable or indistinguishable prior to looking at the claims experience. In other words, we should have no prior knowledge that the tariffs should behave differently. In practice, we most likely know that some tariffs 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 and would drive all prices to the common mean.

4 Application

We will apply the above model to a set of 9 similar outpatient tariffs for women of the DKV, a large German health insurance company. To avoid complications due to the COVID–19 pandemic, we will take the view of the year 2020, i.e. 2020 corresponds to $t = 0$. We will base our estimate on the five years before $T = \{-5, \dots, -1\}$. While we could in principle compare our estimate for $t = 1$ with the real value, it makes no sense because of the pandemic, and it would be impossible in real applications. So we take advantage of the symmetry of our situation and instead of testing the prediction for the second year after the end of our data, we test the prediction for the second year before the start of our data, i.e. for $t = -7$ in our case.

Our model also requires the number of insured, 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 tariffs 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 tariff T9 with only 60 to 100 insured will turn out to be interesting.

Before looking at the observed normalized claims, one must check for outliers that may violate the normality assumption for the 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 tariffs. 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 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 and two years into the past with their 80% prediction intervals.

While this looks stable for several tariffs, for some tariffs 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 0.1% to 7%, which is not plausible for similar tariffs. Both of these observations are exactly the reason why we have introduced our model in the first place.

4.1 Non–Hierachical Models

To proceed with Bayesian modeling, we need to choose priors. We follow the advice of [23] 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

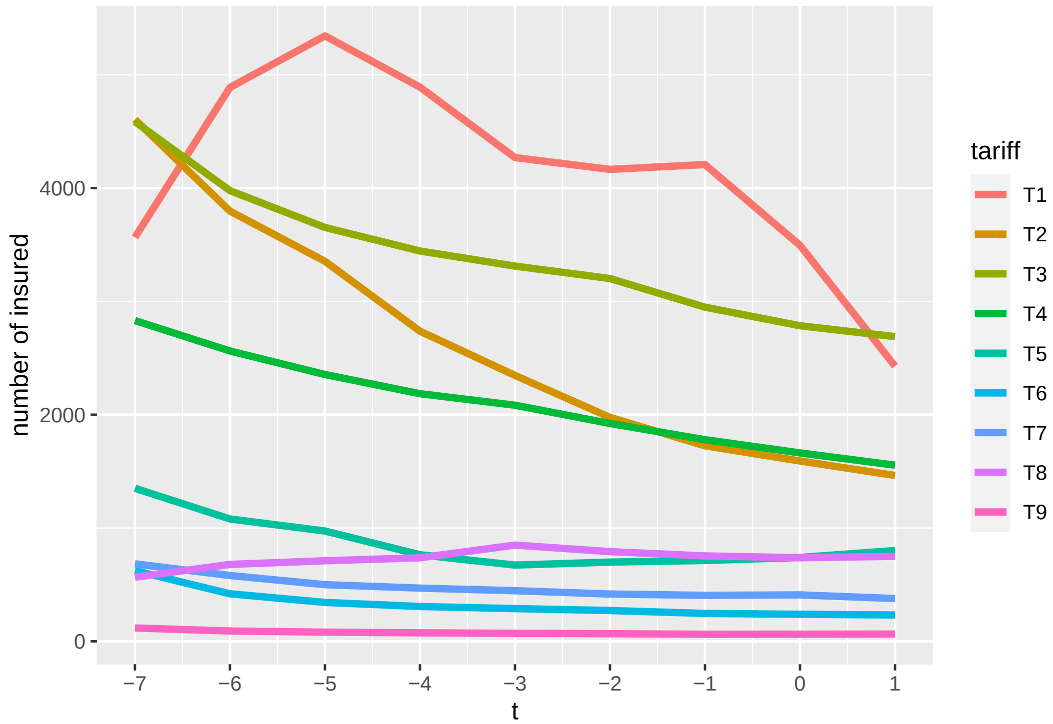


Figure 1: Exposure in outpatient tariffs.

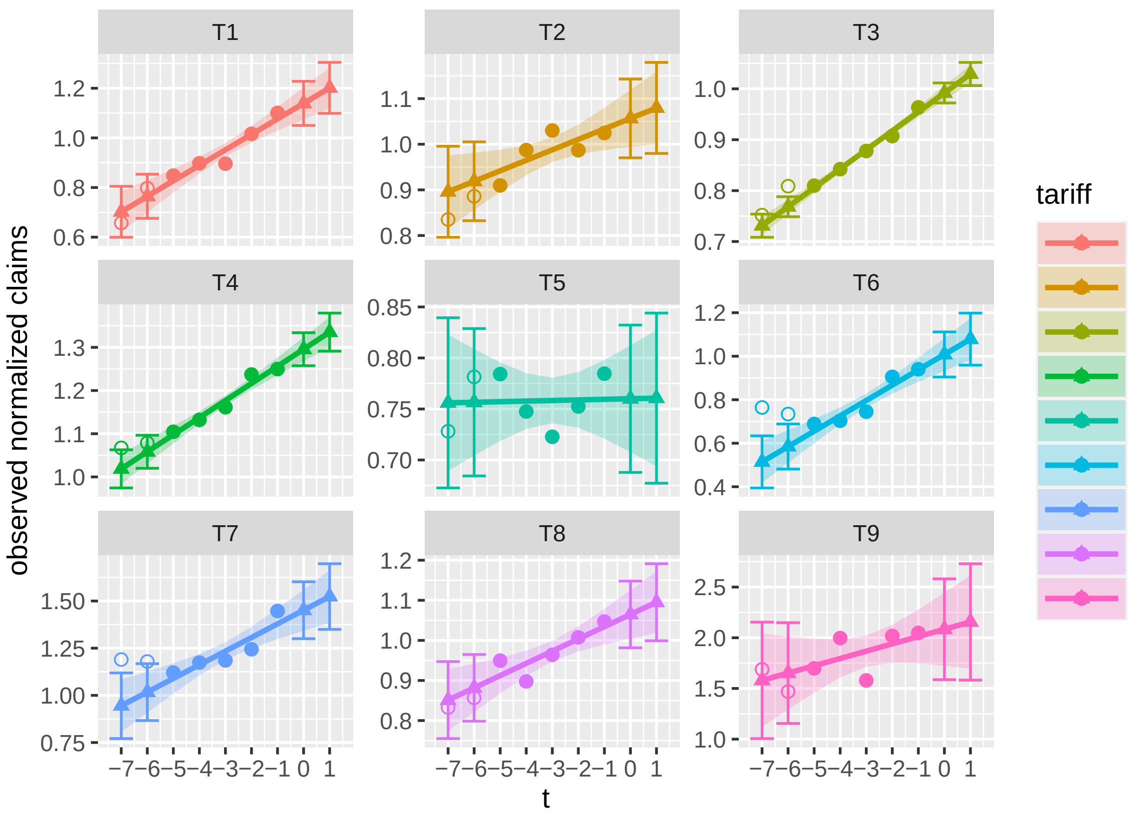


Figure 2: Separately fitted lines for $G^{\theta}(t)$ with 80% confidence intervals. The dots are the observed values used for fitting. The circles are the holdout observations. The triangles are predictions with 80% prediction intervals.

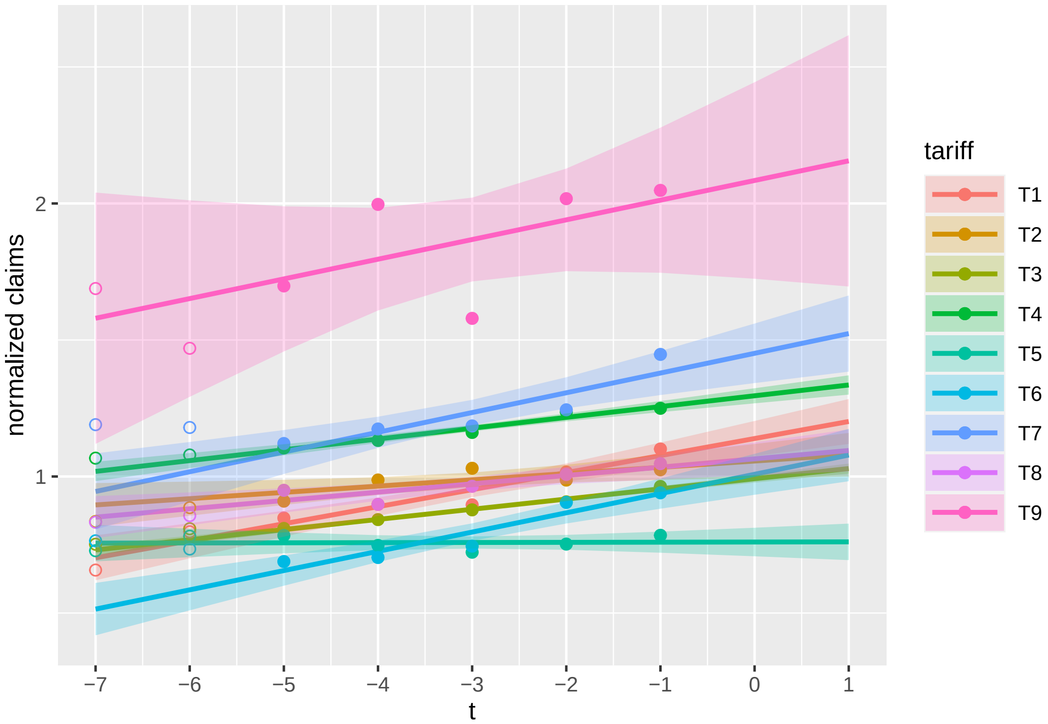


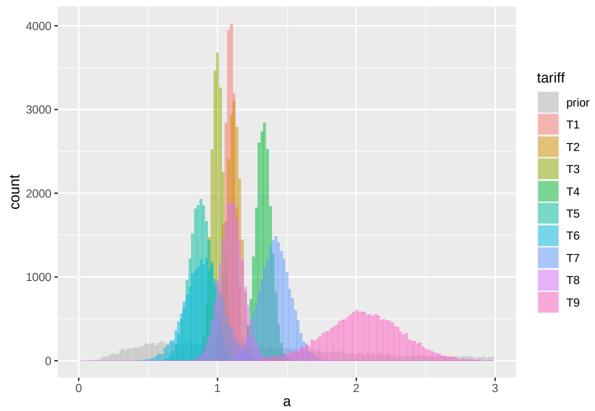
Figure 3: Separately fitted lines for $G^\theta(t)$ with 80% confidence intervals. The dots are the observed values used for fitting. The circles are the holdout observations.

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 by the method described in [23]: 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:

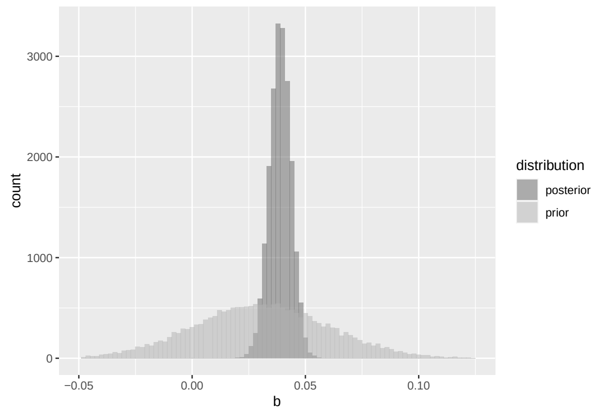
$$\begin{aligned} a^\theta &\sim \text{lognorm}(0.25, 0.75^2) \\ b &\sim \text{normal}(0.03, 0.03^2) \\ \sigma^2 &\sim \text{half–normal}(0, 0.5^2) \end{aligned}$$

The Stan code for the model is given in the Appendix C.1. For code development, the multi–normal distribution examples in the Stan manual [16, Section Gaussian Processes] and the Gaussian processes example by Lemonie [27] were helpful.

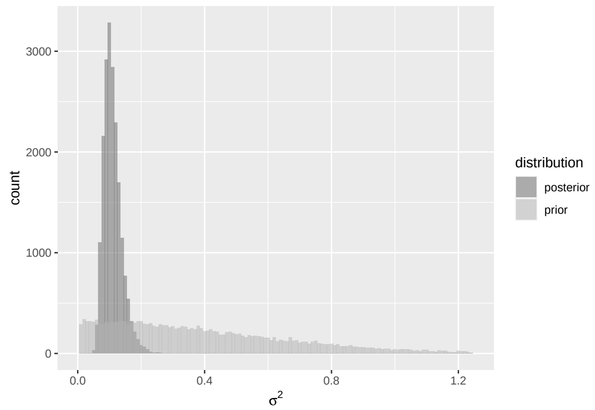
Figure 4d shows the fitted lines for the expected normalized claims together with the observed values, specifically at their test points for $t = -7$ and $t = -6$. As desired, the lines all have the same relative slope and do not feel as random as in Figure 3. In Figure 4b and Figure 4c we see that the 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 tariff θ provides, its distribution is less focused for the smaller tariffs. This is especially the case for the smallest tariff 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. We also note that four of the five points used to fit the line are above the line. Now we have to decide whether we consider the higher claims of the T9 tariff to be probable or not. Here we could confirm that tariff 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.



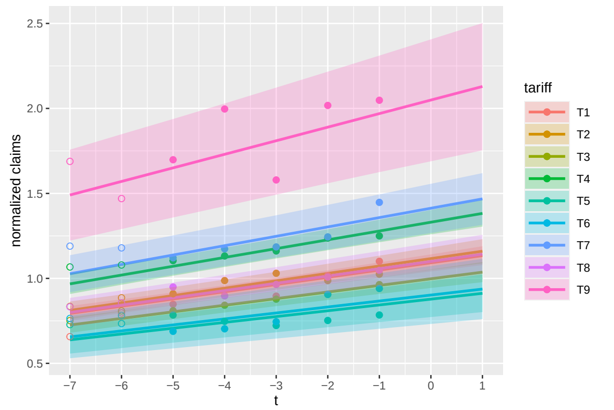
(a) Distribution of a^θ.



(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.

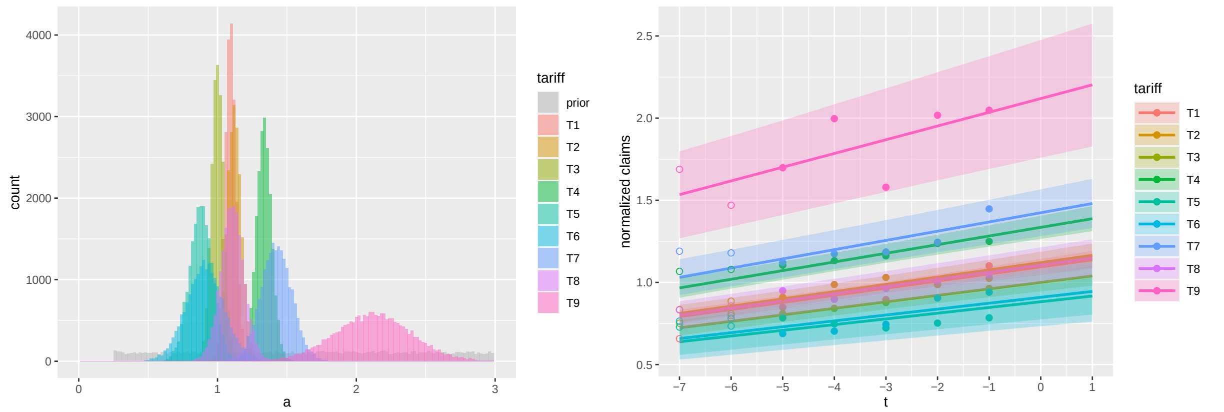
Figure 4: Model: a lognormal b pooled.

We can either exclude the tariff T9 from this set of tariffs 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:

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

In Figure 5a we see the new posteriors for a^θ. The one for tariff T9 has moved to the right, as intended. As intended, the fitted line for tariff 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 small tariffs 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 prior/posterior distributions of our test ${}^{\text{obs}}G^\theta(-7)$ together with the actual realized values as vertical lines. All realized values are quite probable. The one for tariff T1 is the least likely and to the left, the ones for tariff T2 and T3 are near the center, and the rest are to the right side of the distributions. The fact that there are more to the right need not worry us because they belong to the smaller tariffs, so they have less weight. However, the test value for the largest tariff 1 is not very likely, and more importantly, we saw that when we fit the tariffs separately, the tariffs had very different relative slopes. We enforced identical slopes for good reason, but now we want to relax this condition a little and enforce only similar slopes, so that tariff 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.

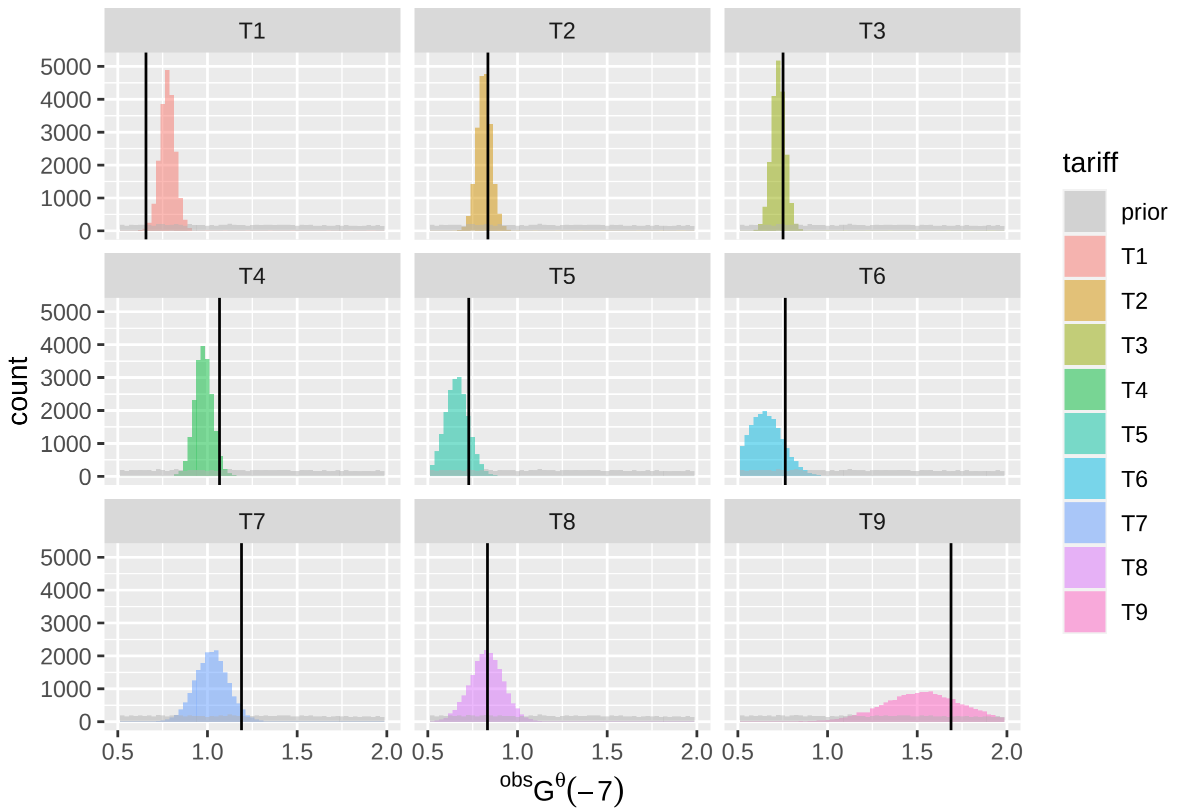


Figure 6: Model: a uniform b pooled: distributions of the test ${}^{\mathrm{obs}}G^\theta(-7)$ with the observed values as vertical lines.

4.2 Hierachical Models

So far, we have assumed that the relative slope b is the same for all tariffs. This was due to our belief that the medical inflation should be the same for tariffs with similar coverage. However, the tariffs 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 tariffs. 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, so 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:

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

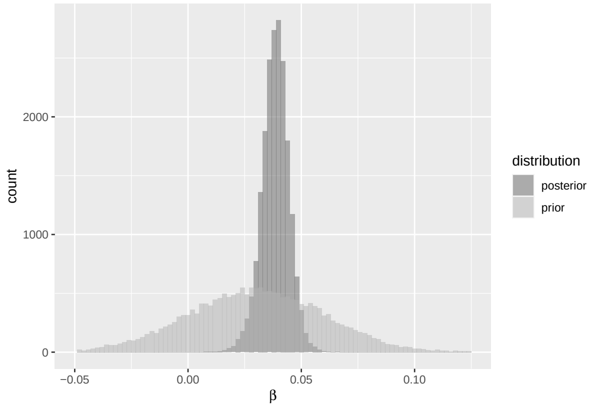
As mentioned at the end of Section 3, hierarchical models can cause computational difficulties, because the full likelihood function can have some numerically unpleasant properties. Here they could be overcome in the usual way by using a non–central parameterization, reducing the step size, and increasing the maximum tree depth. The Stan program for this model is given in Appendix C.2. 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, but that the data implies that the b^θ are even closer together. Their histograms overlap to a large extent, so we show a close–up histogram in Figure 7c. Only the largest tariff T1 provides enough information for its b^θ to be noticeably different. This is not only due to its large size, but also to the significantly above average slope in the individual consideration of the tariffs, see Figure 3 and Table 1. As a result, the test points became more probable, see Figure 8, in particular the one for tariff T1 is less extreme. In Figure 7d we see that the regression lines now intersect due to the different relative slopes. Since this is our favorite model, we also plot them individually in Figure 9. We see that the regression lines lie in the "middle" of the 5 fitting points. However, in contrast to applying the regression separately to tariffs, 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 tariff θ no longer necessarily lie on the regression line of $G^\theta(t)$. Starting from the last observation, the predictions slowly approach the regression line more and more as time progresses, because the last observation error has less and less effect.

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 tariffs, 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:

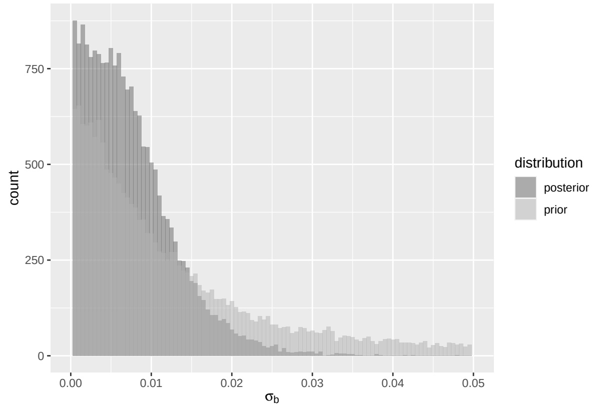
$$
\begin{aligned}
\alpha &\sim \text{lognorm}(0.25, 0.75^2) \\
\sigma_a &\sim \text{half-normal}(0, 0.75^2) \\
a^\theta &\sim \text{lognorm}(\log \alpha, \sigma_a^2) \\
\beta &\sim \text{normal}(0.03, 0.03^2) \\
\sigma_b &\sim \text{half-Cauchy}(0, 0.01) \\
b^\theta &\sim \text{normal}(\beta, \sigma_b^2)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

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.

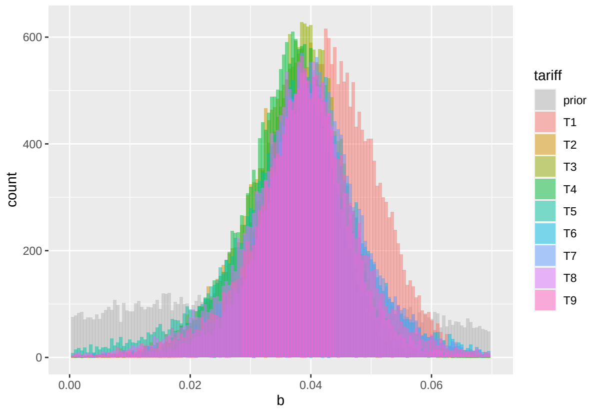
As expected, if we compare the histograms of a^θ of this model variant, Figure 10a, with the histograms of models, in which the a^θ were modeled unpooled, see Figure 4a or Figure 5a, we see that the a^θ have



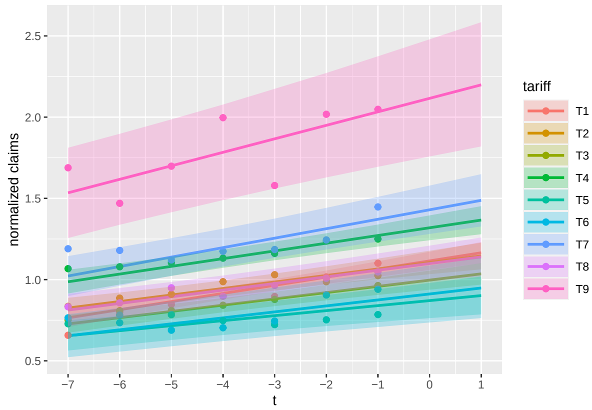
(a) Distribution of β.



(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 10b with Figure 4d or Figure 5b. In this new model variant, we see that a regression line can be completely above or below all observed values of that tariff. 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 tariffs 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 tariffs.

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 tariffs T1, T6, T7. The reason why these tariffs are most affected is that their individual relative inflation deviates most from the common one, so their predictions are most affected by our model choice and therefore most

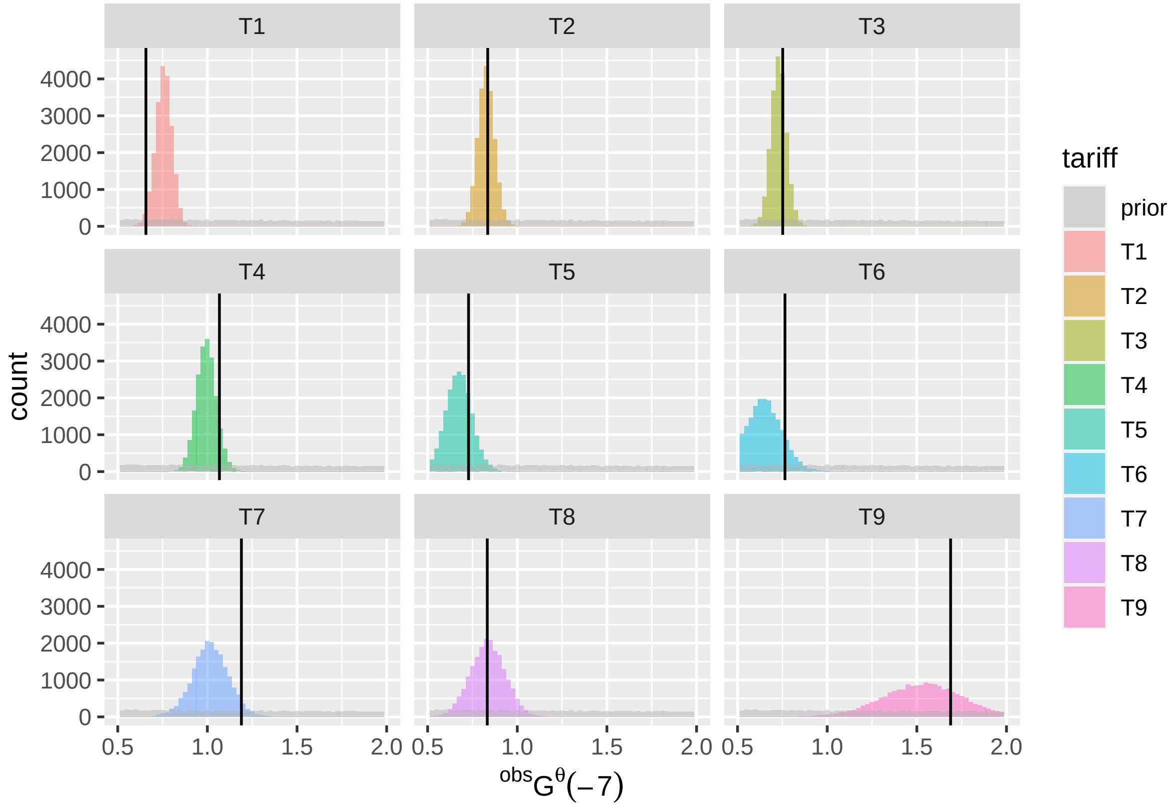


Figure 8: Model: a uniform b hierarchical: distributions of the test $G^{\theta}(-7)$ with the observed values as vertical lines.

affected by the choice of parameters. Of course, the low sensitivity of these parameters is crucial for practical application, since their estimation is difficult and involves some uncertainty, so that in practice one certainly wants to fix them for longer periods of time.

The other sensitivities show that our priors are indeed weakly informative. Changing them affects the predictions by less than 0.5%. This is also true for the σ_b prior, which contains stronger a priori assumptions.

Again, all these effects are dwarfed by our assumption that the tariffs have similar relative inflation, so they are less important. Also, the changes in the predictions in the sensitivities are largest for the tariffs where the individual inflation differs most from the common inflation. Note that our initial belief in similar relative inflation was confirmed by the results of the hierarchical model for b. There we could see that there was little evidence of differences in the relative inflation of the tariffs — only for the largest tariff there was some evidence of higher inflation.

5 Conclusion

We have proposed a model for the claims development of a set of similar tariffs. 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. Only with this assumption a meaningful prediction could be made for most of the smaller, more volatile tariffs; no meaningful trend could be determined from their own data alone. The model relies on a detailed submodel for the observation error of the expected normalized claims. However, 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 tariffs.

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 where we modeled relative

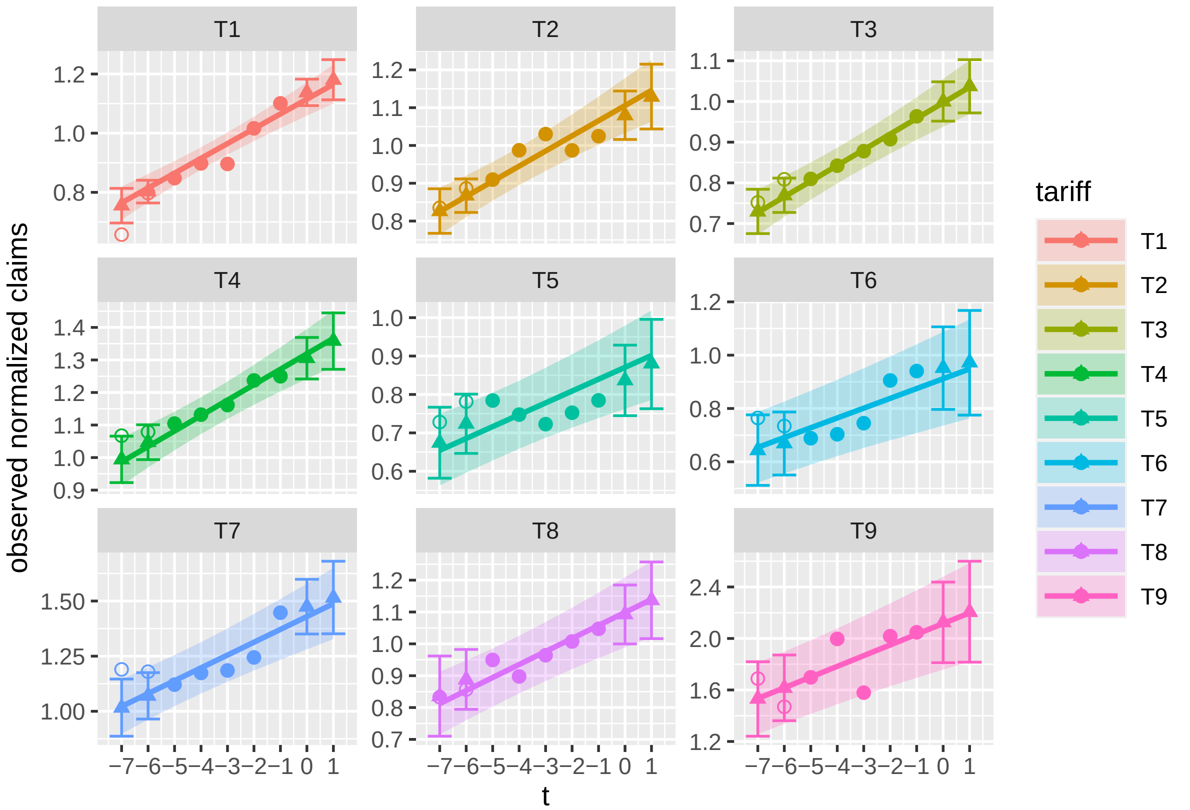
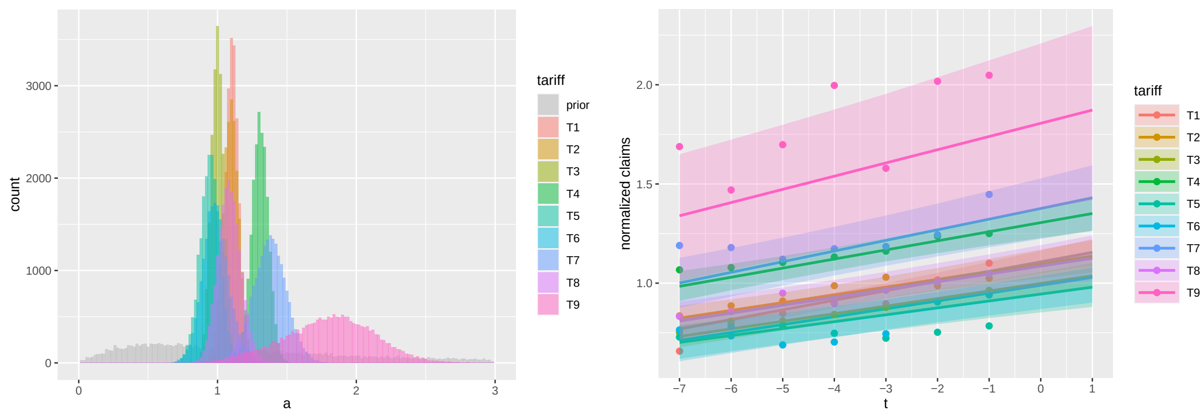


Figure 9: Model: a uniform b hierarchical: fitted lines for $G^{\theta}(t)$ with 80% credibility intervals. The dots are the observed values used for fitting. The circles are the holdout observations. The triangles are predictions with 80% credibility intervals.



(a) Distribution of the a^{θ}.

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

Figure 10: Model: a hierarchical b hierarchical.

inflation hierarchically. There we forced the relative inflations close together, but the data suggested they were even closer together. Only for one tariff there was sufficient evidence of steeper relative inflation. 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 medium–sized tariffs, because only then there can be sufficient evidence for different relative inflation.

Table 1: Sensitivity of the parameter means

model / sensitivity	T1	T2	T3	T4	T5	T6	T7	T8	T9
a^θ									
individual	1.139	1.057	0.992	1.296	0.760	1.008	1.451	1.065	2.084
a lognorm b pooled	1.092	1.116	0.998	1.331	0.878	0.902	1.413	1.100	2.049
a uniform b pooled	1.094	1.121	1.000	1.335	0.882	0.909	1.424	1.106	2.119
a uniform b hierar.	1.115	1.106	0.997	1.319	0.871	0.911	1.430	1.099	2.116
— $\rho+, \lambda+$	1.119	1.107	0.998	1.319	0.881	0.912	1.438	1.098	2.116
— $\iota+$	1.103	1.111	0.996	1.323	0.886	0.894	1.413	1.104	2.110
— σ^2 var+	1.115	1.105	0.998	1.318	0.871	0.914	1.427	1.097	2.113
— β mean+	1.115	1.106	0.997	1.319	0.871	0.911	1.430	1.099	2.116
— β, b^θ var+	1.119	1.101	0.996	1.315	0.866	0.912	1.431	1.095	2.114
b^θ									
individual	0.055	0.022	0.038	0.031	0.001	0.070	0.050	0.029	0.035
a lognorm b pooled	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
a uniform b pooled	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
a uniform b hierar.	0.045	0.036	0.038	0.036	0.035	0.040	0.040	0.037	0.039
— $\rho+, \lambda+$	0.044	0.037	0.039	0.036	0.036	0.040	0.041	0.037	0.039
— $\iota+$	0.044	0.036	0.038	0.035	0.034	0.039	0.040	0.036	0.038
— σ^2 var+	0.045	0.036	0.038	0.036	0.035	0.040	0.040	0.037	0.039
— β mean+	0.045	0.036	0.038	0.036	0.035	0.040	0.040	0.037	0.039
— β, b^θ var+	0.046	0.035	0.038	0.035	0.034	0.040	0.041	0.036	0.038
${}^{\text{obs}}G^\theta(-7)$									
observed	0.657	0.835	0.752	1.067	0.728	0.764	1.190	0.833	1.688
individual	0.702	0.896	0.731	1.019	0.756	0.514	0.945	0.851	1.580
a lognorm b pooled	0.779	0.815	0.728	0.979	0.661	0.647	1.023	0.828	1.517
a uniform b pooled	0.777	0.813	0.726	0.978	0.661	0.647	1.023	0.827	1.533
a uniform b hierar.	0.756	0.826	0.729	0.994	0.674	0.643	1.016	0.835	1.532
— $\rho+, \lambda+$	0.755	0.823	0.729	0.994	0.682	0.639	1.015	0.844	1.532
— $\iota+$	0.759	0.827	0.732	0.998	0.682	0.640	1.016	0.839	1.544
— σ^2 var+	0.756	0.826	0.729	0.995	0.675	0.644	1.015	0.837	1.537
— β mean+	0.756	0.826	0.729	0.994	0.674	0.643	1.016	0.835	1.532
— β, b^θ var+	0.752	0.829	0.730	0.999	0.679	0.641	1.013	0.840	1.539
${}^{\text{obs}}G^\theta(1)$									
individual	1.201	1.080	1.029	1.335	0.761	1.079	1.523	1.095	2.156
a lognorm b pooled	1.157	1.140	1.039	1.372	0.891	0.965	1.503	1.142	2.175
a uniform b pooled	1.159	1.143	1.041	1.376	0.894	0.970	1.510	1.145	2.208
a uniform b hierar.	1.180	1.129	1.037	1.359	0.881	0.973	1.515	1.137	2.205
— $\rho+, \lambda+$	1.186	1.125	1.039	1.357	0.881	0.984	1.530	1.137	2.206
— $\iota+$	1.175	1.129	1.036	1.358	0.886	0.964	1.508	1.138	2.200
— σ^2 var+	1.180	1.129	1.038	1.357	0.881	0.973	1.512	1.135	2.199
— β mean+	1.180	1.129	1.037	1.359	0.881	0.973	1.515	1.137	2.205
— β, b^θ var+	1.184	1.125	1.037	1.354	0.878	0.972	1.517	1.133	2.201

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. Milbrodt H, Röhrs V (2016) Aktuarielle Methoden der deutschen Privaten Krankenversicherung. Verlag Versicherungswirtschaft
4. Gottschalk T, Lax C (2019) Internal documents of the DKV (Deutsche Krankenversicherung)

5. Käsgen H (2023) Stützverfahren für die Berechnung des Grundkopfschadens in der Privaten Krankenversicherung. Bachelor's thesis, Rheinisch–Westfälisch Technische Hochschule Aachen
6. Becker T (2017) Mathematik der privaten Krankenversicherung. Springer Spektrum Wiesbaden
7. Buchner F, Wasem J (2000) Versteilerung der alters- und geschlechtsspezifischen Ausgabenprofile von Krankenversicherern. Zeitschr f d ges Versicherungsw 89:357–392. https://doi.org/10.1007/BF03188231
8. Christiansen M, Denuit M, Lucas N, Schmidt J-P (2018) Projection models for health expenses. Annals of Actuarial Science 12:185–203
9. Piontkowski J (2020) Forecasting health expenses using a functional data model. Annals of Actuarial Science 14:72–82
10. BaFin (2018) Wahrscheinlichkeitstafeln PKV 2017
11. Behne J (1993) Über den wahren und den rechnungsmäßigen Kopfschaden in der Krankheitskosten-Versicherung. Blätter der DGVFM 21:133–140
12. 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.
13. Siegel G (1997) Varianzschätzungen von Kopfschäden. Blätter der DGVFM 23:147--172. https://doi.org/10.1007/BF02808248
14. R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
15. Pinheiro J, Bates D, R Core Team (2023) nlme: Linear and Nonlinear Mixed Effects Models
16. Team SD (2024) Stan Modeling Language Users Guide and Reference Manual, 2.33
17. Bühlmann H, Gisler A (2005) A Course in Credibility Theory and its Applications. Springer–Verlag Berlin Heidelberg
18. Nelder JA, Verrall RJ (1997) Credibility Theory and Generalized Linear Models. ASTIN Bulletin 27:71–82. https://doi.org/10.2143/AST.27.1.563206
19. Frees EW, Young VR, Luo Y (1999) A longitudinal data analysis interpretation of credibility models. Insurance: Mathematics and Economics 24:229–247. https://doi.org/10.1016/S0167-6687(98)00055-9
20. Zuur AF, Ieno EN, Walker N, et al (2009) Mixed Effects Models and Extensions in Ecology with R. Springer–Verlag New York
21. Gelman A, Carlin JB, Stern HS, et al (2013) Bayesian Data Analysis (3rd ed.). Chapman; Hall/CRC
22. 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. https://doi.org/10.1007/s11336-013-9328-2
23. Gelman A, Vehtari A, Simpson D, et al (2020) Bayesian Workflow
24. 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
25. Betancourt M (2017) Diagnosing Biased Inference with Divergences. https://betanalpha.github.io/assets/case_studies/divergences_and_bias.html. Accessed 9 Jan 2024
26. Betancourt M (2020) Hierarchical Modeling. https://betanalpha.github.io/assets/case_studies/hierarchical_modeling.html. Accessed 9 Jan 2024
27. Lemoine N (2017) Even Faster Gaussian Processes in STAN. https://natelemoine.com/even-faster-gaussian-processes-in-stan/. Accessed 20 Mar 2024
28. 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 observed normalized claims is an essential part of any sophisticated normalized 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 tariffs 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 normalized claims ${}^{\mathrm{obs}}G^{\theta}(t)$ are normally distributed with mean $G^{\theta}(t)$. Since ${}^{\mathrm{obs}}G^{\theta}(t)$ is itself a weighted average of $l^{\theta}(t)$ individual claims, it is natural to assume that $\mathrm{Var}({}^{\mathrm{obs}}G^{\theta}(t))$ 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 tariffs, 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 [13, Section 7.3.1], who refers to an older text by Binder that was later published in [28]. In our notation, this would suggest

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

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 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 normalized claim. Second, it is unclear how to deal with lapses and tariff 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 insureds as typical representatives of the tariff. Then we resample the insureds of each tariff 5000 times with replacement, keeping the total number of insureds fixed, i.e. each sample may contain several copies of an insured or none at all. For each sample set of insureds, we take their respective claims and compute the normalized claim as usual. This gives us 5000 realizations of the normalized claim, which we can use to estimate its observation variance.

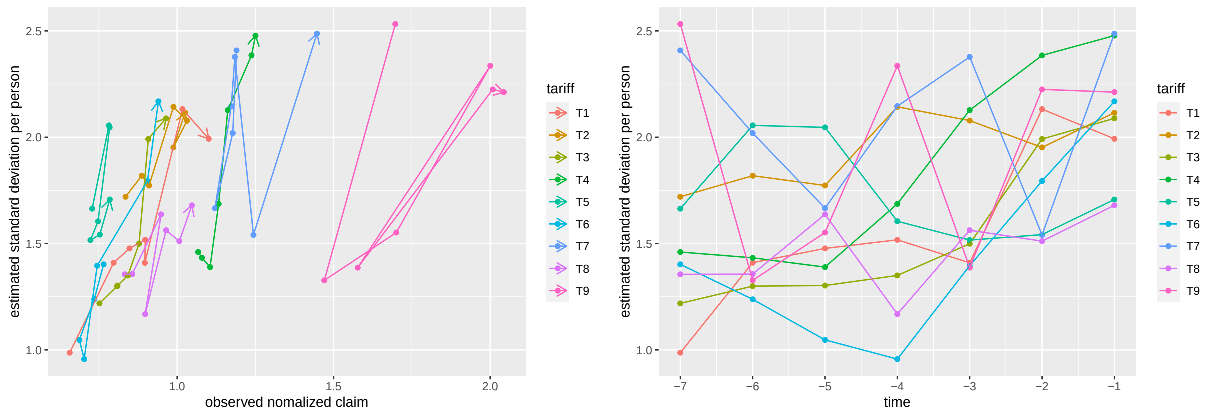
In Figure 11a we plot the "estimated standard deviation per person" $\sqrt{l^{\theta}(t)\hat{\mathrm{Var}}({}^{\mathrm{obs}}G^{\theta}(t))}$ against the ${}^{\mathrm{obs}}G^{\theta}(t)$ connecting successive observations of each tariff. The most striking feature is the apparent randomness, which confirms that the variance estimator is very volatile. There may be some increase in the standard deviation per person as the observed normalized claims increase, but this impression is mostly due to the points in the lower left corner, and these tariffs 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 tariff separately. Here we note an increase in the estimated standard deviation per person over time, see Figure 11b. 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

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

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 tariff θ. Of course, we must remember that we are looking at a set of *similar* tariffs; for non–similar tariffs, the result will surely be different.



(a) the normalized claims connecting the consecutive observations of each tariff.

(b) time.

Figure 11: 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 normalized claim separately by age. Figure 12 shows the result for the four largest tariffs. 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 tariffs 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 tariffs vary by a factor of up to 100, resulting in an exposure effect of up to a factor of 10 between the different tariffs, which clearly dominates the age structure effect.

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 tariff variants (outpatient, inpatient, and dental) to provide a reference. We have chosen tariffs that have a large number of insureds 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 tariff. The tariffs shown here have been on the market for a long time, so the average age of the insured is higher than in the newer tariffs, hence we expect a higher correlation due to a higher prevalence of chronic diseases. This effect is offset by the fact that many insureds changed to newer tariffs, which reduces the correlation in both tariffs. However, we believe that the values presented here are a useful starting point for examining other tariff. For the women in the outpatient insurance, in addition to this tariff here, we also examined the tariffs in the main section and obtained similar values, confirming our belief. A simulation also showed that a medium deductible in the tariff has little effect on the correlations.

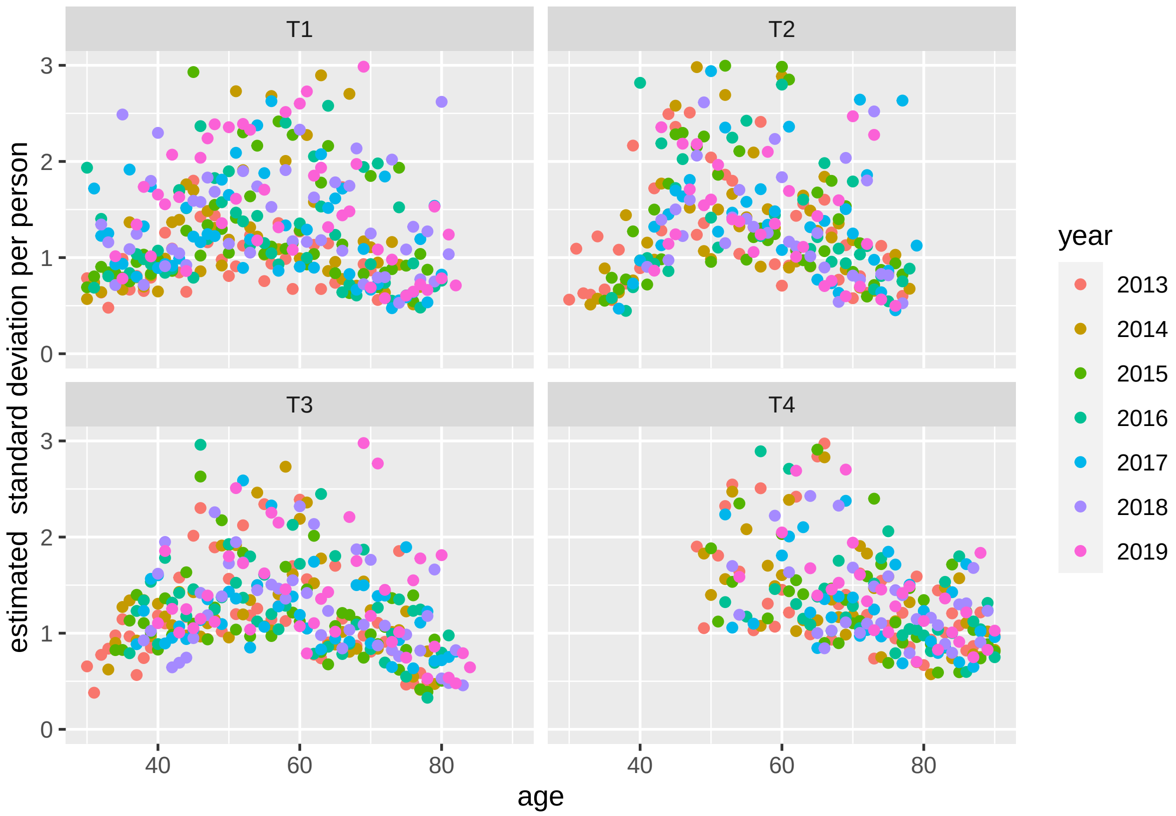


Figure 12: Estimated standard deviation of the normalized claims per person against age.

Unfortunately, outliers significantly affect the correlations. We removed three claims over one million from the inpatient tariffs because there seemed to be a natural gap in claim sizes at around one million. For the outpatient tariffs, 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 13 shows the exposure in the tariffs. They are high enough to expect reliable estimates.

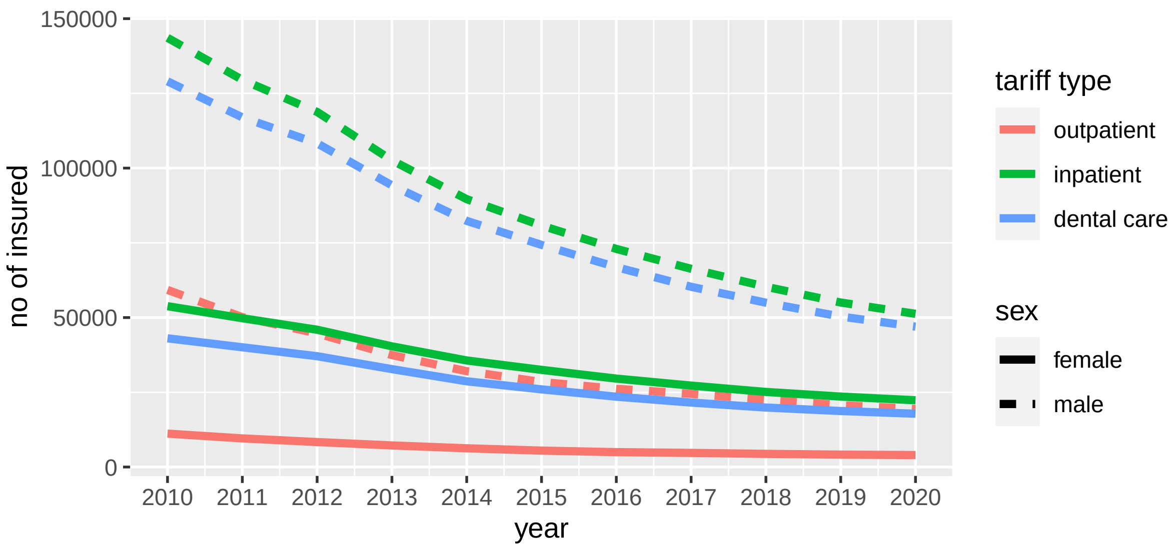


Figure 13: Exposure in the tariffs.

Naturally, we expect that the correlation between two observed normalized claims, $\mathrm{Cor}({}^{\mathrm{obs}}G^{\theta}(t), {}^{\mathrm{obs}}G^{\theta}(t'))$, 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 14 shows the results as a function of t and Δt. We have also computed an average over t by first standardizing the bootstrapped 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 tariff 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 tariffs 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 14 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 14 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.

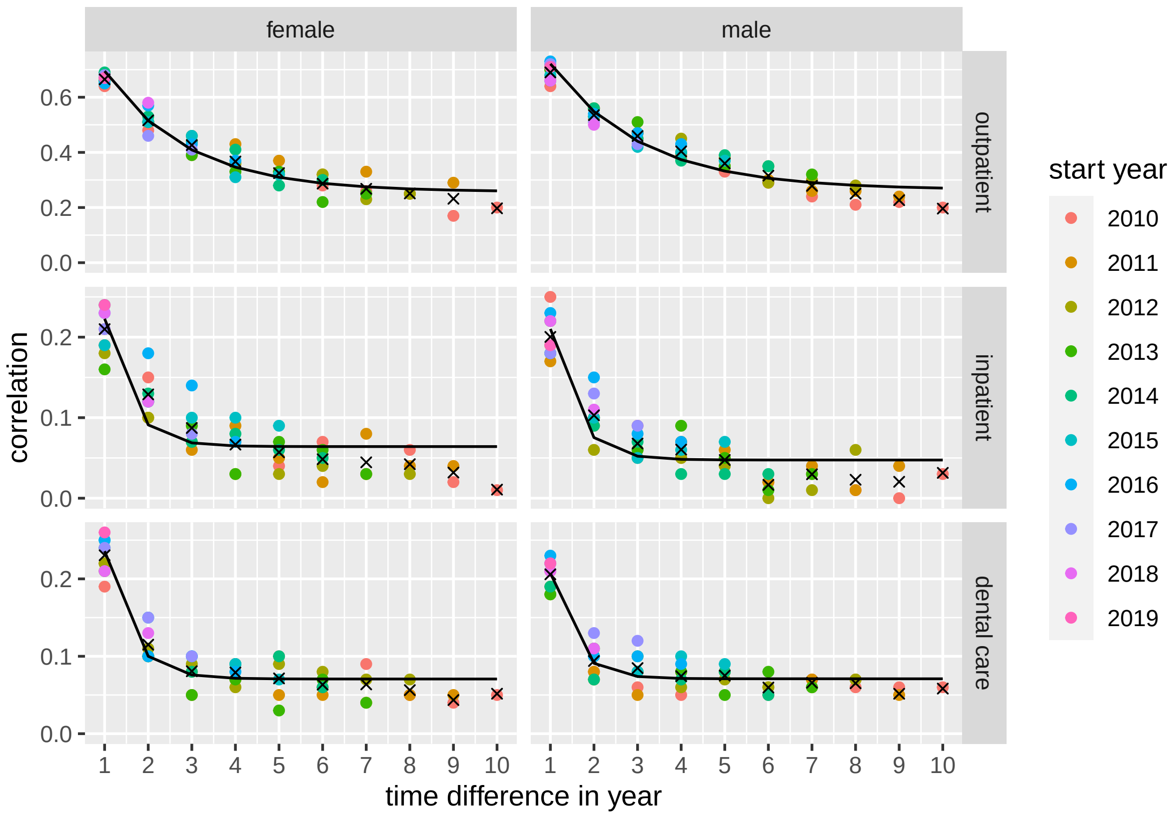


Figure 14: Correlations between years. Crosses are the averages. The lines are the fitted ansatz.

Table 2: Parameters for correlation submodel

tariff 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

C Stan Code

C.1 The a Lognormal b Pooled Model

```
data{
  //dimensions
  int<lower=2> no_years_of_data;
  int<lower=1> no_tariffs;

  // observed data
  vector[no_years_of_data] G_data[no_tariffs]; // claims
  vector[no_years_of_data+4] l[no_tariffs]; // exposure

  // time dependence of Sigma
  real iota;

  // correlation parameters
  real<lower=0, upper=1>  rho;
  real<lower=0, upper=1>  lambda;

  // a lognorm prior parameters
  real a_mu;
  real<lower=0> a_sigma;
  // alternative: for a uniform
  // real<lower=0> a_lower;
  // real<lower=0> a_upper;

  // b normal prior parameters
  real b_mu;
  real<lower=0> b_sigma;

  // sigma^2 observation variance prior parameter
  real<lower=0> sigma2_sigma;
}

transformed data{
  matrix[no_years_of_data+4, no_years_of_data+4] Sigma_fix[no_tariffs];

  // helpers for scaling Sigma
  real min_l_tariff[no_tariffs];
  real min_l;
  for (tariff in 1:no_tariffs)
    min_l_tariff[tariff] = min(l[tariff]);
  min_l = min(min_l_tariff);
```

```
  // construct fixed part of Sigma
  for (tariff in 1:no_tariffs)
    for (i in 1:(no_years_of_data+4)){
      Sigma_fix[tariff][i, i] = min_l / l[tariff][i];
      Sigma_fix[tariff][i, i] *= square(1 + iota * (i-(no_years_of_data+3)));
      for (j in (i+1):(no_years_of_data+4)){
        Sigma_fix[tariff][i, j] = ((1 - lambda) * pow(rho, abs(i-j)) + lambda);
        Sigma_fix[tariff][i, j] *= min_l / sqrt(l[tariff][i] * l[tariff][j]);
        Sigma_fix[tariff][i, j] *= 1 + iota * (i-(no_years_of_data+3));
        Sigma_fix[tariff][i, j] *= 1 + iota * (j-(no_years_of_data+3));
        Sigma_fix[tariff][j, i] = Sigma_fix[tariff][i, j];
      }
    }
}

parameters{
  // a tariff claim level, unpooled
  real<lower=0> a[no_tariffs];

  // b common relative trend
  real b;

  // common observation variance scaling factor
  real<lower=0> sigma2;

  // normalized claims to predict, this year and next
  real G_0[no_tariffs];
  real G_1[no_tariffs];
  // normalized claims to predict in the past for testing
  real G_m0[no_tariffs];
  real G_m1[no_tariffs];
}

transformed parameters{
  // mean vectors
  vector[no_years_of_data+4] mu[no_tariffs];
  for (tariff in 1:no_tariffs)
    for (i in 1:(no_years_of_data+4))
      mu[tariff][i] = a[tariff] * (1 + b * (i-(no_years_of_data+3)));
}

model{
  matrix[no_years_of_data+4, no_years_of_data+4] L_Sigma[no_tariffs];
  vector[no_years_of_data+4] G_obs[no_tariffs];

  // priors
  a ~ lognormal(a_mu, a_sigma);
  // alternative: a ~ uniform(a_lower, a_upper);
  b ~ normal(b_mu, b_sigma);
  sigma2 ~ normal(0, sigma2_sigma);

  // build observation multi-normal distribution
  for (tariff in 1:no_tariffs){
    // append the parameters to the observed ones to predict them
    G_obs[tariff] = append_row(append_row(
                      append_row(G_m1[tariff], append_row(G_m0[tariff],
                        G_data[tariff]
```

```
                    )), G_0[tariff]), G_1[tariff]);

    // Cholesky decompositions of the covariance matrices
    L_Sigma[tariff] = cholesky_decompose(sigma2 * Sigma_fix[tariff]);
  }

  // likelihood
  for (tariff in 1:no_tariffs)
    G_obs[tariff] ~  multi_normal_cholesky(mu[tariff], L_Sigma[tariff]);
}
```

C.2 The a Uniform b Hierarchical Model

```
data{
  // dimensions
  int<lower=2> no_years_of_data;
  int<lower=1> no_tariffs;

  // observed data
  vector[no_years_of_data] G_data[no_tariffs]; // claims
  vector[no_years_of_data+4] l[no_tariffs]; // exposure

  // time dependence of Sigma
  real iota;

  // correlation parameters
  real<lower=0, upper=1> rho;
  real<lower=0, upper=1> lambda;

  // a uniform prior parameters
  real<lower=0> a_lower;
  real<lower=0> a_upper;

  // hierachical b normal prior parameters
  real beta_mu;
  real<lower=0> beta_sigma;
  real<lower=0> b_sigma_gamma;

  // sigma^2 observation variance prior parameter
  real<lower=0> sigma2_sigma;
}

transformed data{
  matrix[no_years_of_data+4, no_years_of_data+4] Sigma_fix[no_tariffs];

  // helpers for scaling Sigma
  real min_l_tariff[no_tariffs];
  real min_l;
  for (tariff in 1:no_tariffs)
    min_l_tariff[tariff] = min(l[tariff]);
  min_l = min(min_l_tariff);

  // construct fixed part of Sigma
  for (tariff in 1:no_tariffs)
    for (i in 1:(no_years_of_data+4)){
      Sigma_fix[tariff][i, i] = min_l / l[tariff][i];
      Sigma_fix[tariff][i, i] *= square(1 + iota * (i-(no_years_of_data+3)));
```

```
      for (j in (i+1):(no_years_of_data+4)){
        Sigma_fix[tariff][i, j] = ((1 - lambda) * pow(rho, abs(i-j)) + lambda);
        Sigma_fix[tariff][i, j] *= min_l / sqrt(l[tariff][i] * l[tariff][j]);
        Sigma_fix[tariff][i, j] *= 1 + iota * (i-(no_years_of_data+3));
        Sigma_fix[tariff][i, j] *= 1 + iota * (j-(no_years_of_data+3));
        Sigma_fix[tariff][j, i] = Sigma_fix[tariff][i, j];
      }
    }
}

parameters{
  // a tariff claim level, unpooled
  real<lower=a_lower, upper=a_upper> a[no_tariffs];

  // b hierarchical relative trend
  real beta;
  real<lower=0> b_sigma;
  real b_tilde[no_tariffs]; // helpers for non-centralized paramterization

  // common observation variance scaling factor
  real<lower=0> sigma2;

  // normalized claims to predict, this year and next
  real G_0[no_tariffs];
  real G_1[no_tariffs];
  // normalized claims to predict in the past for testing
  real G_m0[no_tariffs];
  real G_m1[no_tariffs];
}

transformed parameters{
  // for hierarchical b
  real b[no_tariffs];
  vector[no_years_of_data+4] mu[no_tariffs];

  // non-centralized parameterization: b_tilde must be transformed to b
  for (tariff in 1:no_tariffs){
    b[tariff] = fma(b_sigma, b_tilde[tariff],  beta);
  }

  // mean vectors
  for (tariff in 1:no_tariffs)
    for (i in 1:(no_years_of_data+4))
      mu[tariff][i] = a[tariff] * (1 + b[tariff] * (i-(no_years_of_data+3)));
}

model{
  matrix[no_years_of_data+4, no_years_of_data+4] Sigma[no_tariffs];
  matrix[no_years_of_data+4, no_years_of_data+4] L_Sigma[no_tariffs];
  vector[no_years_of_data+4] G_obs[no_tariffs];

  // a prior
  a ~ uniform(a_lower, a_upper);

  // hierarchical b prior, non-centralized parameterization
  b_sigma ~ cauchy(0,  b_sigma_gamma);
  b_tilde ~ std_normal();
```

```
  // observation variance scaling parameter
  sigma2 ~ normal(0, sigma2_sigma);

  // build observation multi-normal distribution
  for (tariff in 1:no_tariffs){
    // append the parameters to the observed ones to predict them
    G_obs[tariff] = append_row(append_row(
                      append_row(G_m1[tariff], append_row(G_m0[tariff],
                        G_data[tariff]
                    )), G_0[tariff]), G_1[tariff]);

    // Cholesky decompositions of the covariance matrices
    L_Sigma[tariff] = cholesky_decompose(sigma2 * Sigma_fix[tariff]);
  }

  // likelihood
  for (tariff in 1:no_tariffs)
    G_obs[tariff] ~  multi_normal_cholesky(mu[tariff], L_Sigma[tariff]);
}
```