



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER
ACTUARIAL SCIENCE

MASTER'S FINAL WORK
DISSERTATION

COLLECTIVE RISK ASSESSMENT IN AFFORDABLE CARE
ACT MARKETS: A BAYESIAN HIERARCHICAL MODEL

JUAN IGNACIO DE OYARBIDE

OCTOBER 2020



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER
ACTUARIAL SCIENCE

MASTER'S FINAL WORK
DISSERTATION

COLLECTIVE RISK ASSESSMENT IN AFFORDABLE CARE
ACT MARKETS: A BAYESIAN HIERARCHICAL MODEL

JUAN IGNACIO DE OYARBIDE

SUPERVISION:

RUI PAULO

OCTOBER 2020

Acknowledgments

Quiero agradecer enormemente a Eduardo Melinsky, Carolina Castro y Daniel Sarto por guiarme en el comienzo de esta aventura. A Lourdes Centeno por la oportunidad única de permitirme continuar en una institución académica de excelencia. A Rui Paulo por su supervisión y apoyo durante este período de aprendizaje. A la comunidad Stan, y especialmente Asael, por la contribución en este trabajo. A Luís Portugal por su tutoría en mi primera experiencia profesional. A Luis Perez por ayudarme extraordinariamente con todas mis solicitudes. A Matias Pardini por haberme insistido en elegir la Universidad de Lisboa.

A mis hermanos del alma Rodrigue, Arta, Patricia, Stephanie, Andrey, Paula y Lila.

A mi mamá Roxana y tía Patricia por el apoyo incondicional desde el primer día que me fui de casa. Gracias a mis amigos del colegio y de San Lorenzo que están siempre presentes.

Gracias papá, desde donde estés.

Abstract

The changes that the Affordable Care Act introduced to the US health insurance market have entirely altered the traditional ratemaking process. Precisely, the creation of statewide community rating schemes and a guaranteed issue has facilitated insurance coverage to the high-risk population, leading to massive changes in risk pool compositions. The implementation of Risk Adjustment has neutralized some of the consequences of limiting premium variation in the market. However, setting appropriate rate levels has remained cumbersome due to the uncertainty about the statewide risk pool. Many insurers, who could not quantify the health risk associated with the statewide yearly enrollment, had to face unexpectedly high payments on risk equalization. Natsis (2019) stated that in this environment, the use of traditional univariate techniques to project statewide health care costs could be potentially misleading. This thesis proposes a Bayesian approach to reflect important sources of uncertainty over statewide actuarial estimates. The aggregate loss is modeled with a novel collective risk model based on a Generalized Beta Prime (GBP) distribution, accounting for long tail risks and changes in risk pool compositions. The GBP is presented with a mean-dispersion parametrization, which allows the introduction of a hierarchical prior specification over the state-specific means. This parameter structure, responsible of quantifying uncertainty and sharing information among states, is a cornerstone of the adopted collective risk model. Using the Commercial Health Care data extract published by the Society of Actuaries (2019), the model is applied on the *Surgical and Transplant* service category. The resulting heavy-tailed posteriors of the nationwide service means illustrate the high variation of inpatient medical costs. Moreover, the posteriors of the statewide aggregate claims remain highly right-skewed, reflecting the risk of facing sicker populations and high-cost treatments at individual claim level.

Resumo

As alterações que o *Affordable Care Act* introduziu tiveram um impacto significativo no processo de tarifação de seguros de saúde nos Estados Unidos. De forma mais precisa, a criação de um sistema de tarifação regulado, e com cobertura garantida, facilitou o acesso de seguro à população de risco. A inclusão destes indivíduos originou grandes alterações na composição dos grupos de risco de cada estado. A implementação da metodologia do *Risk Adjustment* neutralizou algumas das consequências de restringir as variações de prémios no mercado. No entanto, a estimativa dos níveis de prémios permaneceu complicada devido à incerteza dos riscos coletivos. Muitas seguradoras, que não foram capazes de quantificar corretamente o risco de saúde associado à carteira anual do estado, depararam-se inesperadamente com pagamentos muito altos do *Risk Adjustment*. De facto, Natsis (2019) afirmou que a utilização de técnicas univariadas para projectar os custos médicos neste novo panorama pode produzir resultados enganadores. Nesta tese propomos uma abordagem bayesiana ao problema que pretende incorporar as diversas formas de incerteza presentes em estimativas actuariais ao nível estadual. Implementamos um modelo de risco inovador, baseado na distribuição beta-linha generalizada (BLG), distribuição esta que é capaz de acomodar caudas pesadas e heterogeneidade na composição dos grupos de risco. Apresentamos uma parametrização da distribuição BLG baseada na média e na dispersão, o que permite introduzir uma estrutura paramétrica hierárquica no custo médio. Esta estrutura de parâmetros é a base do modelo para quantificar a incerteza e partilhar informações entre diferentes estados. Utilizando um subconjunto dos dados publicados pela *Society of Actuaries* em 2019, denominados *Commercial Health Care Data*, implementamos o nosso modelo no contexto dos custos associados à categoria *Surgical and Transplant*. Mostramos que a variabilidade nos custos médicos hospitalares de doentes internados conduz a distribuições das médias nacionais *a posteriori* com caudas mais pesadas. Adicionalmente, as distribuições *a posteriori* dos sinistros agregados apresentam um enviesamento à direita muito pronunciado, reflectindo a inclusão no mercado de indivíduos pouco saudáveis e com custos de tratamentos muito altos.

Contents

Acknowledgments	i
Abstract	ii
Resumo	iii
1 Introduction	1
2 Literature review	4
3 Bayesian inference	8
3.1 The Bayesian framework	9
3.2 Hierarchical models	10
3.2.1 Single-level hierarchical model	10
3.2.2 Two-level hierarchical model	12
3.2.3 Borrowing strength property	13
3.3 Markov Chain Monte Carlo methods	15
4 Aggregate loss models	17
4.1 Compound Poisson-Gamma	18
4.2 Hierarchical prior specification	20
4.3 Derivation of the Generalized Beta Prime distribution	21
4.4 Mean-Dispersion parametrization of the Generalized Beta Prime	22
4.5 Individual Risk Model under dependence	23

5	Collective risk assessment in ACA markets	27
5.1	ACA Risk Adjustment	28
5.2	Commercial Health Care Data Extract 2009-2015	29
5.2.1	Exploratory data analysis	30
5.2.2	Collective Risk Model: Compound Negative Binomial - Generalized Beta Prime	39
5.2.3	Prior distributions	42
5.2.4	Posterior predictive sampling	44
5.3	Numerical results	44
6	Conclusions	49
A	Generalized Beta Prime	51
A.1	GBP: Compound Gamma-Gamma	51
B	CHC dataset	52
B.1	Actuarial indicators	52
B.2	Model outputs	54
B.2.1	Small group	56
B.2.2	Individual	60
C	Stan code	65

Chapter 1

Introduction

Health insurance markets are one of the most regulated in the US insurance industry and they are of central interest in public policy. While governments must guarantee accessibility and affordability of health services, insurers must deal with financial risks carried by increasing unpredictability of health conditions. In general, a grand part of the losses is incurred by a small proportion of insureds who represent the highest health risk, e.g., chronic patients. Furthermore, the costs associated with this group have great variation among insurers and populations.

Over the past decades, governments have created community rating schemes to generate cross subsidies between the low-risk and high-risk individuals. Necessarily, these schemes have been supported by risk equalization programs to avoid market instability (Neuhaus, 1995); some examples are the risk structure compensation (RSC) in Germany, Risk Equalization Fund (REF) in the Netherlands, and the Risk Equalisation Trust Fund (RETF) in Australia. In the US, the Affordable Care Act (ACA) has imposed a modified community rating scheme and a guaranteed issue in the small group and individual market segments since 2014. These rules imply that insurers can only use a predefined set of rating factors for pricing, and they can no longer deny coverage. In order to mitigate the financial impact of high-risk enrollees moving between insurers, the federal government created Risk Adjustment.

Traditionally, actuaries would focus on the enrollee population and individual risk profiles for pricing. However, under the ACA, insurance rates must reflect the statewide population health status, given that the risk associated to health conditions is spread to the market pool and equalized among insurers through Risk Adjustment. In this context, sicker statewide populations may represent a higher utilization of health care resources, driving insurance rates up. Since its implementation, Risk Adjustment has been increasingly important in the insurers' revenues. Two examples are the closures of Northwell Health CareConnect plans in New York (Livingston, 2017) and HealthyCT in Connecticut (Zorn, 2016), both with significant payments made to the market pool.

Generally, one can represent the statewide health care cost with two principal stochastic components, the frequency of health care utilization and the cost per medical service. Natsis (2019) in collaboration with the Society of Actuaries provided an extensive univariate analysis for the small group and individual market segments from the year 2009 to 2015. The small group segment, related to group health insurance, has shown a relatively stable frequency and average cost along the years. However, individual markets were massively altered by the ACA changes introduced in 2014. These changes allowed a large part of the previously uninsured high-risk population access the health insurance system, leading to unequal changes in statewide risk compositions. This new mix of risks not only increased the demand on health care resources but also shifted it to particular medical services.

Furthermore, the costs associated with specific inpatient admissions or treatments based on new specialty drugs have increased long tail risks for insurers, e.g., admissions due to hemophillia cost 0.15M USD on average (Chen, 2016) and the drug Zolgensma costs 2.125M USD per patient per year (Rosenberg, 2019), among others. In this context, the assessment of statewide pools can be a difficult task, primarily because of the different sources of uncertainty involved in the loss generating process. In fact, Natsis (2019) stated that, in highly uncertain markets, the use of univariate techniques to project health care costs can be inaccurate and potentially misleading.

In this thesis, we propose a full Bayesian analysis of statewide collective risks in ACA small group and individual markets. We consider previous collective risk models presented by Migon and Moura (2005), Migon and Penna (2006) and Amin and Salem (2015). However, in our approach, we propose a novel distributional representation of the conditional aggregate claims, in order to capture the long tail risks, on the one hand, and the yearly changes on the statewide pool, on the other. From the Bayesian standpoint, we assume a hierarchical Gamma prior on a rate parameter associated to the statewide pool. Moreover, we derive the unconditional 3-parameters distribution of the individual claims known as Generalized Beta Prime (GBP). Indeed, this marginalization allows us to compress an original two-level hierarchical model into a single-level model and then reparametrize the unconditional target distribution as a function of its mean and dispersion parameters, similarly to the structure of a Generalized Linear Model (Ohlsson and Johansson, 2010). The latter is of particular interest given that we rebuild a hierarchical structure over the state-specific mean parameters, introducing shrinkage effects among risk pools and recovering a full probabilistic representation of nationwide costs per service.

Using the Commercial Health Care data extract published by the Society of Actuaries (2019), we apply our model on the *Surgical and Transplant* medical service

breakout, for the small group and individual market segments. We show that the increasing uncertainty in the individual market segment in 2015 is fully reflected on the posterior distributions of the nationwide cost per service, resulting in wider credible intervals compared with the small group segment. However, the uncertainty on per-member-per-month (PMPM) claim cost posterior predictive distributions varies significantly among statewide markets in the two segments. It is noteworthy that this uncertainty is generated from unobservable quantities that are fully specified in the hierarchical parameter structure.

This thesis is organized as follows. Chapter 2 presents a literature review about the Bayesian approach in Actuarial Science and the existing work on Bayesian collective risk assessment in health insurance. Chapter 3 unfolds the Bayesian framework, hierarchical modeling, and sampling methods. Chapter 4 reviews aggregate loss models and the development of the novel GBP model. Chapter 5 is devoted to the ACA Risk Adjustment and the application of the model previously developed. Lastly, Chapter 6 contains the main conclusions and other final thoughts.

Chapter 2

Literature review

Bayesian methodologies have been implemented in insurance since at least the beginning of the past century. Particularly, in the context of credibility theory, it was Whitney (1918) who introduced, with the use of inverse probabilities, the general credibility formula that is widely known today. In his paper, Whitney pursued the problem of finding the posterior distribution of the real hazard in a workers' compensation insurance contract, i.e., the probability of death, given the collective and individual experiences. In his words

The problem of experience rating arises out of the necessity, from the standpoint of equity to the individual risk, of striking a balance between class-experience on the one hand and risk experience on the other.

His motivation relied on how to improve individual estimates by using all the information available; therefore, incorporating the observations from other insurance contracts that belong to the same class. Under a Bayesian approach, he assumed that the number of deaths in a single contract follows a Binomial distribution, where the probability of occurrence follows itself a Normal distribution. *A priori*, this parameter, which represented the contract-specific probability of death, was centered at the class-specific mean. He showed that the resulting individual estimates could be written as weighted averages between the observed contract-specific and group proportions.

At the time, Whitney was not able to provide full posterior distributions. However, he showed how posteriors would look like if the number of deaths was assumed to follow a Normal distribution, a case where the posterior and the prior are conjugate distributions. Besides the practical difficulties for numerical integration, this paper settled the foundation of Bayesian hierarchical modeling in the context of credibility theory, illustrating how we can borrow information from similar risks through prior specification.

A few decades later, Bailey (1950) consolidated the interpretation and notation of posterior distributions as such in the context of credibility, referring to Laplace’s Generalization of Bayes’ rule. He extended the Binomial-Normal model proposed by Whitney to a Binomial-Beta model, and additionally studied the Poisson-Gamma model. He also introduced linear estimates obtained with a least squares regression and explored the results for conjugate priors. This field, known as Linear Bayes, was formally formulated two decades later by Bühlmann and Straub (1970), with successive development; see, for example Hachemeister (1975), Venter (1985), De Vylder (1984), Neuhaus (1984), Goldstein and Wooff (2007), among others.

Following Bailey’s work, Mayerson (1964) reinforced the philosophical interpretation of the Bayesian paradigm as an updating mechanism of beliefs, e.g., considering the manual rate *a priori* knowledge. In the Linear Bayes field, he restated previous results of conjugate distributions, named as exact credibility. A decade later, Jewell (1974) generalized these results to the exponential family of distributions.

Lindley and Smith (1972) presented a Bayesian hierarchical linear model within a modern framework. They proposed the current definition of *hyperparameters* and *hyperpriors* focusing on parameter dependencies. They showed in which circumstances Bayes estimates can perform better than ordinary least squares in terms of mean squared error. Additionally, they compared the resulting partial-pooling effect with a ridge regression.

Jewell (1975) brings this set-up to credibility theory with his paper “The Use of Collateral Data in Credibility Theory: A Hierarchical Model”. He proved that the Bühlmann and Straub (1970) model is a special case of hierarchical modeling (two-level model with “diffuse” priors). Moreover, he asserted that hierarchical models allow individual risks’ estimates to borrow collateral information from others that belong to the same cohort; hence, generating partial-pooling at observational level while collapsing manual rate information at the second level, following also Taylor (1974). Another significant point, in Jewell’s words, was the following

Thus, in a hierarchical model, we hope to use nationwide statistics, together with all the data from our portfolio, not only to predict next year’s fair premium for individual risks, but also to draw inferences about what kind of a portfolio we have.

It is noteworthy that the work presented in the seventies and eighties goes in line with a shift of the Bayesian philosophical standpoint as an updating mechanism of beliefs towards a broad modeling architecture viewed from an hypothetico-deductive perspective; see, e.g., Gelman and Shalizi (2013). For instance, Lindley and Smith

(1972) made a comparison of hierarchical models with ridge regression, which relates prior distributions, or *hyperpriors*, with regularizing parameters. Furthermore, Jewell (1975) expressed that prior information is actually not needed for defining manual rates but a well-specified model configuration that utilizes all the collateral data available.

Panjer and Willmot (1983) made a point about Bayesian uncertainty models in collective risk theory, particularly in modeling claim frequency. Meyers and Schenker (1983) expressed their standpoint about the sources of uncertainty in large insured groups. Interestingly, they stated

The traditional models used in collective risk theory, such as the generalized Poisson distribution, do not allow for uncertainty in estimating the expected loss. This may be acceptable for the small insured, since the variance of the losses due to the random nature of the loss process is large compared to the variance due to the misestimation of the expected loss. As the insured increases in size, however, the variance due to the misestimation of the expected loss will dominate.

Since the eighties, Bayesian hierarchical modeling has been viewed not only as a framework to cross-inform individual estimates but also reflect different sources of uncertainties on the representation of observable quantities; therefore, capturing complex features in the data that traditional models were not able to handle. In parallel, with the development of advanced sampling algorithms, especially in the nineties with Markov Chain Monte Carlo (MCMC), full Bayesian analysis became plausible for high-dimensional model structures. Some of these sampling algorithms can be found in Geman and Geman (1984), Gelfand and Smith (1990), Gelman et al. (1992) and Tierney (1994), among others.

Klugman (1991) presented a comprehensive overview of hierarchical modeling. He reviewed some of the shortfalls of the linear approximation in experience rating, encouraging actuaries to apply a full Bayesian approach. In addition, he presented model assessment techniques such as predictive checks. Gelman et al. (2013) presented a modern Bayesian workflow, consolidating a formal framework to carry on a data analysis process.

Precisely, Gelman et al. (2015) introduced Stan: a probabilistic language to specify full Bayesian models. In order to draw samples from the posterior distribution, Stan utilizes the No-U-Turn-Sampler (NUTS), a variation of Hamiltonian Monte Carlo (HMC). For the past years, this algorithm has been widely used in several disciplines, especially to sample from the posterior in multi-dimensional models; an introduction

can be found in McElreath (2020) and Betancourt (2017). In the actuarial field, Stan has been used to implement statistical models only by a few authors; see, for example, Gao et al. (2018), and Gesmann and Morris (2020).

In the context of health insurance, Migon and Moura (2005), Migon and Penna (2006) and Amin and Salem (2015) presented different variations of Bayesian hierarchical collective risk models. Their final goal was to obtain full posterior distributions of the aggregate losses in a health care plan, segmenting by age band and time period. Particularly, a common aspect is that individual claim amounts are modeled as conditionally independent and identical Gamma distributions, while the dependence is introduced through a hierarchical Gamma prior on a single rate parameter. In the context of ACA markets, we find this model structure inadequate to capture yearly changes in risk composition and increasing long tail risks. On the other hand, the partial pooling is not fully interpretable given that it is performed over the rate and shape parameters of the assumed Gamma distribution.

In this thesis, motivated by the aforementioned points, we assign a single hierarchical parameter to the yearly statewide population and, sequentially, we marginalize out such hyperparameter in the individual claim distribution. This procedure leads to a 3-parameter distribution known as Generalized Beta Prime or Pearson type VI; see e.g., Venter (1983) and Kupper (1962). This marginalization allows us to, first, hold the variation introduced by the hierarchical parameter and, second, to propose a new parametrization based on the mean and a dispersion parameter, similar to the structure of a Generalized Linear Model (Ohlsson and Johansson, 2010).

Over the past decades, several authors have studied Beta regression models with a mean parametrization; see, e.g., Ferrari and Cribari-Neto (2004) and Grün et al. (2011). However, to our knowledge, this set-up has not yet been explored for the Generalized Beta Prime. Indeed, we show that our parametrization not only improves the chain trajectories of the HMC-NUTS sampler but also lays the ground for an interpretable hierarchical structure to cross-inform individual-specific means, make inference on group mean costs, and tackle exposure imbalance in the data.

Lastly, similarly to Sarabia et al. (2016), we show that the sum of dependent but identically Generalized Beta Prime distributed claim amounts, with the dependence following a specific pattern, is also Generalized Beta Prime distributed. Moreover, we implement this distribution in a collective risk model that we use to process the Commercial Health Care data (Society of Actuaries, 2019). We obtain relevant information from the resulting posteriors and, ultimately, we present a framework that can be extensively used for risk analysis at the present time.

Chapter 3

Bayesian inference

In parametric statistics, we build probabilistic models to describe the behavior of observable quantities, generated by a stochastic phenomenon occurring in the world, and make inference about unobservable quantities of interest (i.e., model parameters). Normally, we assume a distributional assumption for the observables, while parameter inference and model predictions differ according to the approach to statistics. Two important schools are the frequentist, which connects probability to the frequency of events in large samples, and the Bayesian, whose foundation lays on the use of probability to quantify uncertainty, whether on observable or unobservable unknowns (McElreath, 2020).

Under the frequentist approach, we rely on experimental design and sampling distributions to make inference, and point estimation methods (e.g., maximum likelihood) to perform model predictions. On the other hand, the Bayesian paradigm is based on the application of Bayes' theorem to infer about model parameters, considered as unknown random quantities. Moreover, model predictions are represented with a full probability distribution that reflects parameter uncertainty and is conditional on the observed data; this topic will be discussed in the following section.

In this thesis, we approach the problem of estimating health care costs in ACA markets within a Bayesian framework. In Chapter 4, we argue that the Bayesian approach is particularly suited for this problem because of the highly uncertain insurance environment, missing or partial information and hierarchically structured data. Interestingly, one can achieve a complex but meaningful model structure by using a hierarchical prior specification (Lindley, 1975); this topic will be covered in Section 3.2. Furthermore, the fit of these models would have not been possible without the exponential gain on computational power in the past years, alongside the development of advanced sampling algorithms. We conclude this chapter with a brief introduction to these methods.

3.1 The Bayesian framework

Considering a parameter vector θ and observable data $y = (y_1, \dots, y_n)$, one can write the posterior distribution of θ using Laplace generalization of Bayes' theorem

$$\pi(\theta|y) = \frac{L(\theta|y)\pi(\theta)}{L(y)}, \quad (3.1)$$

where $L(\theta|y)$ is a likelihood function of parameters, $\pi(\theta)$ is the prior distribution of θ and $L(y)$ is the marginal likelihood or probability of the data y . Moreover, since the latter is a normalizing constant, one can write the posterior distribution up to a constant of proportionality as

$$\pi(\theta|y) \propto L(\theta|y) \times \pi(\theta). \quad (3.2)$$

Generally, the form of $L(\theta|y)$ depends on a probabilistic assumption made for the observable quantities y . On the other hand, the prior $\pi(\theta)$ represents our uncertainty about θ before having observed the data. In practice, there usually are no single choices for $L(\theta|y)$ and $\pi(\theta)$, hence the importance of assessing their adequacy along the model-building process (Gelman et al., 2013).

The selection of the prior has been the main criticism of the Bayesian approach for several decades (Efron, 1986). In this thesis, we merely use priors to construct a meaningful parameter structure that we use to learn from the data, make inference on unknown quantities of interest (e.g., nationwide means) and reflect parameter uncertainty on the stochastic representation of a new observation, as introduced next.

Suppose that we are interested in predicting a new observation \tilde{y} , that is conditionally independent of y given θ . Assume that we obtained the posterior distribution $\pi(\theta|y)$, then we can write the posterior predictive of \tilde{y} as

$$f(\tilde{y}|y) = \int f(\tilde{y}|\theta)\pi(\theta|y) d\theta, \quad (3.3)$$

where $f(\tilde{y}|\theta)$ is the conditional mass (density) function of a new observation. The resulting probability distribution can be used not only to provide model predictions (point estimates), but also to communicate the uncertainty over these, a critical source of information for decision-making (Berger et al., 2006).

In practice, there exist several actuarial problems that are governed by parameter uncertainty (Meyers and Schenker, 1983). In fact, in ACA markets, the changes of enrolment population and the development of new medical treatments significantly

increases uncertainty in the estimation process. Therefore, the implementation of traditional models and consequent statistical inference can be potentially misleading. As explained in the following section, a hierarchical prior specification can capture the missing random components, hence affecting the resulting posterior (3.2). This uncertainty is passed through the model structure, and is ultimately reflected on model predictions, which are represented by a full probability distribution (3.3).

3.2 Hierarchical models

Hierarchical models have gained great importance in recent years due to their ability to represent complex data structures. These models have shown a significant improvement over traditional statistical models in terms of predictive accuracy (Gelman, 2006). Generally, one can build hierarchical models by specifying prior distributions that are conditional on new parameters, or *hyperparameters*, which have their own prior distributions, or *hyperpriors*. Then, one can proceed with the application of Bayes' theorem as usual and recover the posterior distribution of all parameters involved.

This section starts with a simple single-level hierarchical model and continues with a two-level hierarchical model, both presented in the context of an actuarial application. To conclude, we introduce the borrowing strength property and the link with credibility theory.

3.2.1 Single-level hierarchical model

We consider next the actuarial problem of estimating the expected claim cost of a specific insurance coverage in different areas of a country. Suppose that individual claim amounts are not observed and we only have summary statistics of the data. This scenario is typical in insurance given the regulation for data protection. In fact, the application of this thesis, presented in Chapter 5, is made on statewide aggregate data and individual claims are not provided.

Let $\bar{y} = (\bar{y}_1, \dots, \bar{y}_J)$ be a vector of conditionally independent observations that represent the average claim cost for area $j = 1, \dots, J$. We can propose the following simple model

$$\begin{aligned}\mu_j &\sim N(a, b), \\ \bar{y}_j | \mu_j &\sim N(\mu_j, \sigma_j^2), \quad j = 1, \dots, J,\end{aligned}$$

where $\sigma_j^2 = \sigma^2/n_j$, being σ^2 a known variance parameter and n_j the total observed number of claims in area j , and a, b fixed hyperparameters for the prior $\pi(\mu_j)$.

Figure 3.1 shows a directed acyclic graph (DAG) of the single-level hierarchical model presented here. DAGs are graphical representations that illustrate the relationships between the quantities involved in a statistical model. They play an essential role in hierarchical models due to the high number of parameters usually considered. Notice that one can visualize in Figure 3.1 two levels of model parameters, however the second level is composed by fixed quantities a, b, σ, n_j .

The posterior of the parameter vector $\mu = (\mu_1, \dots, \mu_J)$ can be written, up to a constant of proportionality, as

$$\pi(\mu|y) \propto L(\mu|y) \times \pi(\mu), \quad (3.4)$$

where $L(\mu|y)$ is the likelihood function and $\pi(\mu)$ the joint prior. Given the independence between parameters, the latter can be denoted as

$$\pi(\mu) = \prod_{j=1}^J \pi(\mu_j). \quad (3.5)$$

Now consider a new observation \tilde{y}_j for area j that is conditionally independent of y given μ . We can denote the posterior predictive of \tilde{y}_j as

$$f(\tilde{y}_j|y) = \int f(\tilde{y}_j|\mu)\pi(\mu|y) d\mu, \quad (3.6)$$

where $f(\tilde{y}_j|\mu)$ is the Normal density conditional on the parameter vector and $\pi(\mu|y)$

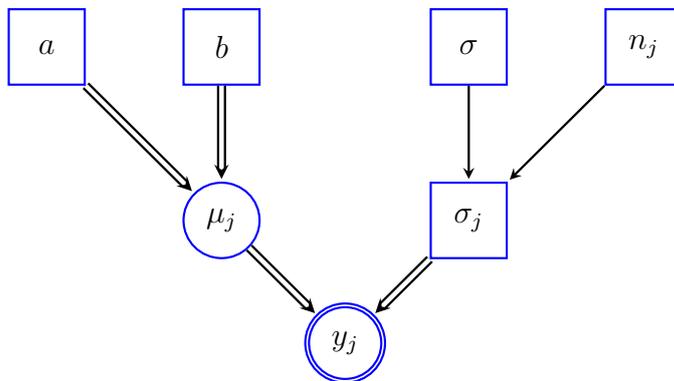


Figure 3.1: Single-level hierarchical model DAG. The nodes represent quantities of interest. Square symbols represent known quantities and circles represent stochastic quantities. Single-arrows describe functional relationship, while double-arrows denote stochastic dependence. Double contour lines indicate observable quantities.

is the posterior given the observed data y .

Expression (3.6) reflects our uncertainty on model parameters, contained in the posterior $\pi(\mu|y)$, over the density of a new observation \tilde{y}_j . In this case, $f(\tilde{y}_j|y)$ is only affected by the observations from area j , given the *a priori* independence between group-specific means (3.5). Generally, point estimates obtained under a single-level model are known as *unpooled* estimates.

3.2.2 Two-level hierarchical model

Suppose now that we relax the assumption of fixed hyperparameters in the priors $\pi(\mu_j)$ of the previous example. Instead, we consider that individual means μ_j are assumed to follow Normal priors centered at a common hyperparameter ω with known variance σ_μ^2 , for $j = 1, \dots, J$. Then, the model can be specified as follows

$$\begin{aligned}\omega &\sim N(c, d), \\ \mu_j|\omega &\sim N(\omega, \sigma_\mu^2), \\ \bar{y}_j|\mu_j &\sim N(\mu_j, \sigma_y^2),\end{aligned}$$

where the vector $\mu = (\mu_1, \dots, \mu_J)$ is now conditional on the hyperparameter ω assumed to follow a Normal with fixed hyperparameters c, d . Figure 3.2 illustrates the DAG corresponding to this two-level hierarchical model.

The resulting posterior, up to a constant of proportionality, is

$$\pi(\mu, \omega|y) \propto L(\mu, \omega|y) \times \pi(\mu, \omega), \quad (3.7)$$

where $\mu = (\mu_1, \dots, \mu_J)$ is a parameter vector of the individual means, $L(\mu, \omega|y)$ the likelihood function and $\pi(\mu, \omega)$ the joint prior.

In order to visualize the parameter structure, one can write the joint prior as a chain of dependencies (Kruschke, 2014)

$$\pi(\mu, \omega) = \pi(\mu|\omega)\pi(\omega), \quad (3.8)$$

where $\pi(\mu|\omega)$ is the prior distribution of the first-level parameter vector μ conditional on the second-level hyperparameter ω , whose hyperprior is denoted as $\pi(\omega)$.

Now consider a new observation \tilde{y}_j for area j that is conditionally independent of

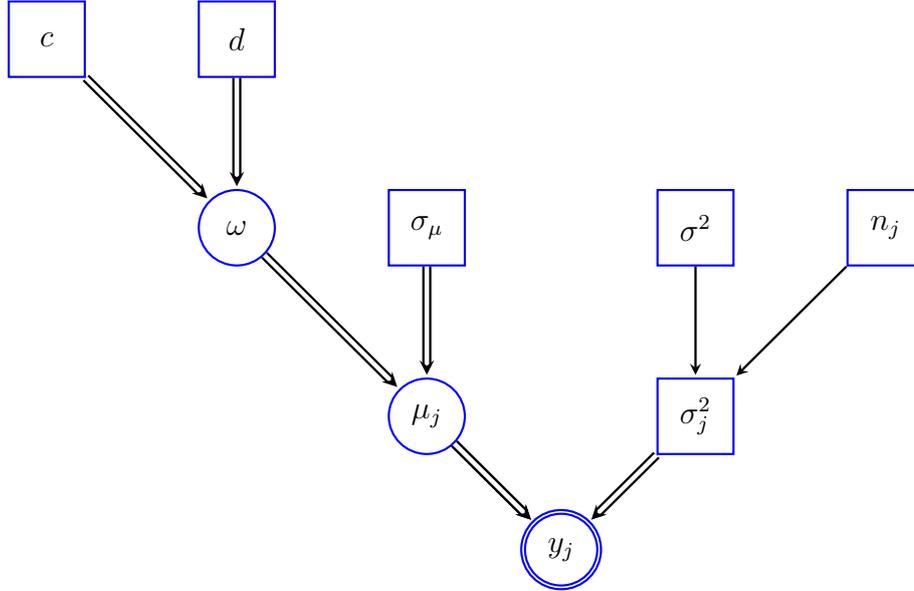


Figure 3.2: Two-level hierarchical model DAG. The nodes represent quantities of interest. Square symbols represent known quantities and circles represent stochastic quantities. Single-arrows describe functional relationship, while double-arrows denote stochastic dependence. Double contour lines indicate observable quantities.

y given μ and ω . Under this model, we can denote the posterior predictive for \tilde{y}_j as

$$f(\tilde{y}_j|y) = \int \int f(\tilde{y}_j|\mu, \omega) \pi(\mu, \omega|y) d\omega d\mu, \quad (3.9)$$

where $f(\tilde{y}_j|\mu, \omega)$ is the Normal density conditional on the parameter vector and $\pi(\mu, \omega|y)$ is the joint posterior given the observed data y .

The probability density function (3.9) reflects our uncertainty about model parameters over a new observation; therefore, passing the uncertainty from the overall group mean ω and the group-specific mean μ_j . In this case, $f(\tilde{y}_j|y)$ is affected by the observations from every area $j = 1, \dots, J$, given the *a priori* dependence among group-specific means introduced by ω (3.8). This information sharing across groups, known as the borrowing strength property, is presented in the following section.

3.2.3 Borrowing strength property

Gelman et al. (2007) implemented the two-level hierarchical model presented in the previous section, except that they assumed an improper prior distribution for ω , that is a prior proportional to 1 (Gelman et al., 2013). They showed that the posterior means of μ_j in this case, conditional on the variance parameters σ_y^2 and σ_μ^2 , can be written as

$$\hat{\mu}_j = \frac{(n_j/\sigma_y^2)\bar{y}_j + (1/\sigma_\mu^2)\hat{\omega}}{n_j/\sigma_y^2 + 1/\sigma_\mu^2}, \quad (3.10)$$

where

$$\hat{\omega} = \frac{\sum_{j=1}^J \bar{y}_j / (\sigma_y^2/n_j + \sigma_\mu^2)}{\sum_{j=1}^J 1 / (\sigma_y^2/n_j + \sigma_\mu^2)}. \quad (3.11)$$

In the context of credibility theory, Jewell (1975) presented 3.10 in the following form

$$\hat{\mu}_j = \frac{n_j}{n_j + \sigma_y^2/\sigma_\mu^2} \bar{y}_j + \left(1 - \frac{n_j}{n_j + \sigma_y^2/\sigma_\mu^2}\right) \hat{\omega}, \quad (3.12)$$

that is

$$\hat{\mu}_j = Z \bar{y}_j + (1 - Z) \hat{\omega}, \quad (3.13)$$

where Z represents the credibility factor and $\hat{\omega}$ is the manual rate obtained from all the data available. As observed in these expressions, posterior means are weighted averages between group-specific averages and the resulting overall mean from all the groups. This shrinkage effect towards the overall mean $\hat{\omega}$ is known as partial-pooling or borrowing strength property.

Particularly, one can see in Expression 3.12 that when the number of observations for group j increases, the credibility to the data \bar{y}_j increases as well. Moreover, a higher variance among means reduces the factor σ_y^2/σ_μ^2 , resulting in an increase of the credibility to the group-specific data as well. In contrast, a high variance σ_y^2 reduces the credibility factor Z , consequently assigning more weight to the overall mean. The exact amount of shrinkage will ultimately depend on how the hierarchical model is parametrized, priors on variance parameters (σ_y^2 and σ_μ^2), the number of observations in each group, and the actual variation in the data.

Gelman (2014) discusses the importance of using strong priors in high hierarchical levels. The reason is founded on the fact that we rarely have very precise information at observational level but often have it at a higher hierarchical level, which can significantly impact the analysis. For instance, in insurance we might not have prior information about the distribution of individual losses in a small book of business. However, we might expect that the group-specific mean does not significantly differ from the overall mean, given similar underwriting conditions. This variation across sub-populations can be explicitly incorporated in a Bayesian hierarchical model.

In the previous section, we showed how a two-level hierarchical model reflects parameter uncertainty of a higher-level parameter, ω , over model predictions (3.9). Moreover, in this section we presented how this model structure allows to borrow information from all the data available, resulting in group-specific means that are

weighted averages (3.13). Hierarchical modeling is, therefore, a flexible technique to combine several quantities in a meaningful structure (see e.g., Figure 5.11, page 42), capture uncertainty and improve model predictions (Lindley, 1975). However, many decades had to pass before the computational tools, necessary to implement them in full generality, were developed. In fact, it was the advent of Markov Chain Monte Carlo (MCMC) that eventually made this possible. These algorithms are introduced in the following section.

3.3 Markov Chain Monte Carlo methods

One of the main challenges of the Bayesian approach is the computation of the marginal likelihood, i.e., the denominator in Bayes' rule (3.1). In general, this integral does not have a closed-form expression, so exact posterior inference is not feasible. One solution to handle this problem is to draw samples from the posterior distribution in order to get a simulation-based representation of this probability (Kruschke, 2014). There are several methods that have been implemented to achieve this goal. Particularly, we introduce in this section: Markov Chain Monte Carlo (MCMC) methods, which includes Hamiltonian Monte Carlo (HMC) and the algorithm used by Stan: No U-Turn Sampler (NUTS) (Gelman et al., 2015); an extensive introduction to MCMC methods can be found in Robert and Casella (2013).

Markov Chain Monte Carlo (MCMC) is an algorithm that constructs a Markov chain, and hence a sequence of correlated draws, to stochastically explore a particular target distribution. In the long run, the frequencies of the drawn values converge to the probabilities under the target distribution, i.e., the stationary distribution is the distribution from which it is desired to obtain samples. In Bayesian inference, the target is the joint posterior distribution and the state space of the Markov chain corresponds to the parameter space of our statistical model.

Let $\{\theta_t\}_{t \geq 0}$ be a discrete-time Markov chain on a continuous state space Θ , where θ represents an unknown parameter in a statistical model. The following equality holds, as long as $E(g(\theta)|y)$ exists

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t g(\theta_i) = E(g(\theta)|y). \quad (3.14)$$

Thus, expectations with respect to the posterior can be calculated by time averaging the function of interest over realizations from a single chain trajectory. This facilitates the calculation of posterior predictive distributions and other expectations over the

parameter space (Smith and Roberts, 1993).

Metropolis et al. (1953) first implemented an MCMC method to approximate intractable integrals in physics, which, a few decades later, was generalized by Hastings (1970). Briefly, the Metropolis algorithm consists of two steps: first, a parameter value is proposed using a candidate-generating density, which defines the rule of generating value proposals in every transition; and second, the action of transitioning to the proposed value is evaluated according to an acceptance probability, constructed to guarantee reversibility.

The principal challenge of Hastings (1970) is the definition of the candidate-generating density. This definition has an impact on the rejection rate and ultimately on the convergence rate to the stationary distribution. If the proposals are made too close to the current position, the chain may take a long time to fully explore the support of the target distribution. On the contrary, if proposals are made far away, the rejection rate will be high and the chain will not move smoothly, given that at each rejection the chain stays in the current position.

Hamiltonian Monte Carlo (HMC), an algorithm that was developed by Duane et al. (1987), proposes a different approach of making proposals. HMC is based on the construction of a bivariate distribution of the model parameter and an auxiliary variable called momentum. From this distribution, we obtain a system of differential equations (Hamiltonian equations) whose approximate solution is used to generate candidate parameter values. One solution to this system is the *leapfrog* integrator, which conserves the volume and guarantees ergodicity (Betancourt, 2017). Basically, this integrator is used to make steps across level sets of the bivariate distribution for a given exploration time (tuning parameter). Then, it returns a new momentum and a candidate parameter value, which is evaluated according to an acceptance-rejection step as in Hastings (1970).

Therefore, HMC consists on two steps: one deterministic, that depends on how long the Hamiltonian trajectory is integrated, and one stochastic, following an acceptance probability. The novelty feature of HMC is that it performs the exploration by learning about the shape of the target distribution. This is indeed very convenient when the volume of the parameter space becomes considerably large, as in high-dimensional hierarchical models, and the algorithm can make proposals in a more efficient manner (Betancourt, 2017). However, one of the drawbacks of HMC is the sophistication required for tuning the auxiliary parameters. If the integration time is long, the exploration might return to the original place and the chain will not move smoothly. In addition, if the step size of the *leapfrog* integrator is large, the approximation to the equations, and hence the exploration, will be inaccurate. On the other

hand, when the step size is small the method becomes computationally expensive and the convergence might be slow.

The No-U-Turn-Sampler (NUTS), recently developed by Hoffman and Gelman (2014), brings a solution to the drawbacks of HMC. In fact, it proposes a dynamic way of calibrating the auxiliary parameters, avoiding inefficient exploration of Hamiltonian trajectories. NUTS can automatically adapt the number of *leapfrog* steps in each transition so the chain can move as far as possible without returning to its original position. This facilitates the exploration in posteriors that present complex curvatures without the need of re-calibrating the auxiliary parameters. In this thesis, we use this algorithm, which is implemented by Stan, to sample in our collective risk model. Then, we mainly focus on the convergence and efficiency of the chains, e.g., with the assessment of trace plots (Figure B.1 in Appendix B.2).

In previous sections, we explored the Bayesian approach and how hierarchical models represent parameter uncertainty over the posterior predictive distribution (3.9). This distribution provides information about the plausibility of future outcomes given a parameter structure and the observed data. Furthermore, in this section, we presented advanced sampling algorithms that made the fit of complex high-dimensional models possible. Next, we provide a theoretical background on aggregate loss models, a widely explored field in Actuarial Science. Then, using the concepts presented here and, motivated by the context of the ACA, we present our contributions to the collective risk theory.

Chapter 4

Aggregate loss models

The study of aggregate loss models has been extensively developed in the literature; e.g., Panjer (1980), Kaas et al. (2008), Klugman et al. (2012), Dickson (2016). Precisely, these models are centered on the aggregate loss random variable (i.e., the total claim amount incurred by a group of policyholders within an exposure period). This quantity is fundamental for the insurer to define risk management policies (e.g., reinsurance), estimate actuarial reserves or calculate capital requirements, among others. The purpose of this chapter is to build an aggregate loss model associated with statewide risk pools, and specifically, to certain medical claims where, even at

state level, the total number might not be significant (e.g., inpatient admissions due to *Musculoskeletal* conditions). The Bayesian framework provides a model-building approach, through hierarchical prior identification, to capture important sources of uncertainty in the different components of the aggregate loss model.

We start this chapter by presenting the collective risk model, that is, the aggregate loss represented as the sum of N claim amounts where N is itself a random variable (Klugman et al., 2012). Particularly, we first introduce the traditional compound Poisson model (Embrechts et al., 2013). Migon and Moura (2005) developed a hierarchical parameter structure in this model to capture extra variation and share information across different risk groups in a health care plan. This chapter shows how, by further developing this structure, we can derive the novel Generalized Beta Prime model, which is indeed a generalization of other models already explored in the literature.

4.1 Compound Poisson-Gamma

The collective risk model studies the aggregate loss random variable $S = X_1 + \dots + X_N$ by separately modeling the number of claims N and the individual claim amount X_j , for $j = 1, \dots, N$. One of the advantages of assuming a separated model for the number of claims (or frequencies) and the claim amounts is that we can make inferences about quantities of interest in each of the components. Then, the uncertainty about these quantities is ultimately reflected on the posterior predictive of the aggregate loss. In this section, we first introduce the Bayesian approach with a simple Compound Poisson model, when only the aggregate claims and number of claims are observed.

Consider $s = (s_1, \dots, s_m)$ conditionally independent observations of aggregate claims and $n = (n_1, \dots, n_m)$ conditionally independent claim counts, where each pair (s_i, n_i) represents different outcomes from a risk pool. Then, one can express the model as follows

$$N_i | \lambda \sim \text{Poisson}(\lambda), \quad (4.1)$$

$$S_i | N_i = n_i, \alpha, \beta \sim \text{Gamma}(n_i \alpha, \beta), \quad (4.2)$$

where $i = 1, \dots, m$ is the observation identifier, $S_i | N = n, \alpha, \beta$ represents the conditional aggregate claims, $N_i | \lambda$ the number of claims, λ the expected number of claims, α a shape parameter and β a rate parameter. Furthermore, the number of claims is usually associated with risk exposure, e.g., the number of member months in the pool. For now, we leave the exposure out of the analysis.

The joint density of the conditional aggregate claims and claim counts can be expressed as

$$f(s_i, n_i | \lambda, \alpha, \beta) = \frac{\beta^{n_i \alpha} s_i^{n_i \alpha - 1} e^{-\beta s_i} \lambda^{n_i} e^{-\lambda}}{\Gamma(n_i \alpha) n_i!}. \quad (4.3)$$

Thus, the likelihood function can be written as

$$L(\lambda, \alpha, \beta | s, n) \propto \prod_{i=1}^m \frac{\beta^{n_i \alpha} s_i^{n_i \alpha - 1} e^{-\beta s_i}}{\Gamma(n_i \alpha)} \lambda^{n_i} e^{-\lambda}. \quad (4.4)$$

and the posterior

$$\pi(\lambda, \alpha, \beta | s, n) \propto \prod_{i=1}^m \frac{\beta^{n_i \alpha} s_i^{n_i \alpha - 1} e^{-\beta s_i}}{\Gamma(n_i \alpha)} \lambda^{n_i} e^{-\lambda} \pi(\lambda, \alpha, \beta), \quad (4.5)$$

where $\pi(\lambda, \alpha, \beta)$ represents the joint prior density.

With the implementation of a sampling algorithm on (4.5), one can easily recover samples from the marginal posterior densities $\pi(\lambda | s, n)$, $\pi(\alpha | s, n)$ and $\pi(\beta | s, n)$. Moreover, the aggregate claims posterior predictive for a new observation \tilde{s} , that is conditionally independent of s and n given λ , α and β , is denoted as

$$f(\tilde{s} | s, n) = \sum_{\tilde{n}=0}^{\infty} \left[\int \int \int f(\tilde{s} | \tilde{n}, \alpha, \beta) f(\tilde{n} | \lambda) \pi(\lambda, \alpha, \beta | s, n) d\lambda d\alpha d\beta \right], \quad (4.6)$$

where $f(\tilde{s} | n, \alpha, \beta)$ is the conditional Gamma density, $f(\tilde{n} | \lambda)$ the probability mass function of the claim counts model, which so far has been assumed to be Poisson distributed, and $\pi(\lambda, \alpha, \beta | s, n)$ the posterior density of model parameters.

An MCMC sample with I iterations can be represented in a three-columns matrix, where each column represents a parameter, and each row an iteration. In this case,

$$H_{I,3} = \begin{pmatrix} \lambda^{(1)} & \alpha^{(1)} & \beta^{(1)} \\ \lambda^{(2)} & \alpha^{(2)} & \beta^{(2)} \\ \vdots & \vdots & \vdots \\ \lambda^{(I)} & \alpha^{(I)} & \beta^{(I)} \end{pmatrix}.$$

In order to obtain a sample from the aggregate claims posterior predictive, one should draw from $N \sim \text{Poisson}(\lambda^{(i)})$ and successively $S | N = n \sim \text{Gamma}(n\alpha^{(i)}, \beta^{(i)})$ for $i = 1, \dots, I$.

The single-level hierarchical model presented here (as shown in Section 3.2.1)

might lead in some contexts to a misrepresentation of the aggregate loss' variation and tail (Meyers and Schenker, 1983). Migon and Moura (2005) developed a hierarchical prior specification for the conditional aggregate claims to fully reflect cost uncertainty in a small health care plan. We present next their development and, for the purpose of the section, we consider the number of claims as a known quantity.

4.2 Hierarchical prior specification

One characteristic of the Compound Poisson model presented in the previous section is that all aggregate claims, for $i = 1, \dots, m$, share the same value of the parameters α and β . Using the concepts of hierarchical modeling, Migon and Moura (2005) relaxed such assumption to capture extra variation and share information across risk classes. Thus, assuming that β_i and α_i are drawn from Gamma hyperpriors with common hyperparameters, one can express the model as follows

$$\alpha_i | \kappa, \zeta \sim \text{Gamma}(\kappa, \zeta), \quad (4.7)$$

$$\beta_i | \gamma, \theta \sim \text{Gamma}(\gamma, \theta), \quad (4.8)$$

$$S_i | n_i, \alpha_i, \beta_i \sim \text{Gamma}(n_i \alpha_i, \beta_i), \quad (4.9)$$

where n_i is the number of claims, α_i is a shape parameter now Gamma distributed with hyperparameters κ and ζ , and β_i a rate parameter also Gamma distributed with hyperparameters γ and θ . Notice that under this model structure S_i 's are now dependent. Implicitly, this model structure implies that the j^{th} individual claim amount for group i follows a Gamma distribution $X_{ij} | \alpha_i, \beta_i \sim \text{Gamma}(\alpha_i, \beta_i)$, for $j = 1, \dots, n_i$. Figure 4.1 shows the relationship among parameters for this hierarchical structure. This model is a special case of the individual risk model, that is, when n is a known quantity and the individual components are identically distributed (Klugman et al., 2012).

The hierarchical structure presented here augments the single-level model shown in the previous section to a two-level model. Although this parameter structure is reasonable for a small group health care plan, it is not enough to capture important sources of uncertainty introduced by the ACA in the US health insurance market (e.g., changes in yearly enrolment). Moreover, the Gamma distributional assumption for the individual claim amounts is not appropriate due to its light tail (Venter, 1983). It is the purpose of the next section to propose a reasonable distribution for medical insurance claims and a meaningful hierarchical structure that can capture parameter uncertainty.

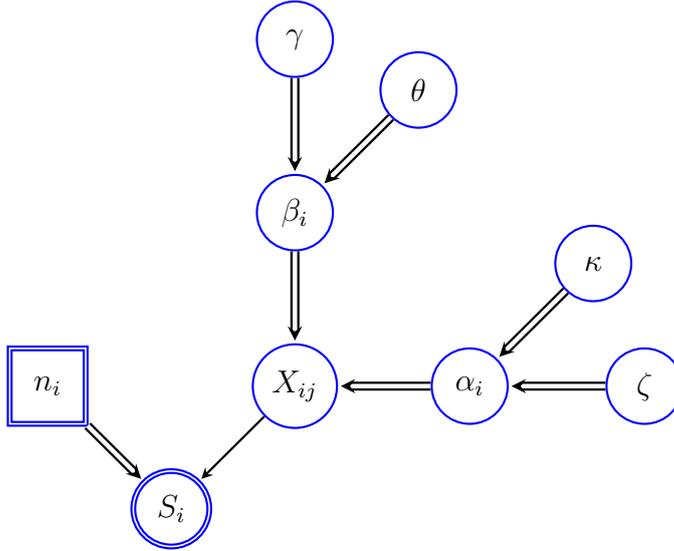


Figure 4.1: DAG Individual Risk Model with Gamma distributed individual components. The nodes represent quantities of interest. Square symbols represent known quantities and circles represent stochastic quantities. Single-arrows describe functional relationship, while double-arrows denote stochastic dependence. Double contour lines indicate observable quantities.

4.3 Derivation of the Generalized Beta Prime distribution

Over the past years, insurers participating in the US health insurance market have been exposed to long tail risks, e.g., inpatient admissions due to hemophilia can cost 0.15M USD per patient per year (Chen, 2016), specialty drugs such as Spinraza 0.375M USD and Zolgensma 2.125M USD (Rosenberg, 2019). Furthermore, the inclusion of previously uninsured population provoked massive changes in risk pool compositions (KFF, 2019). These dynamics have generated more uncertainty on aggregate claims that cannot be fully captured by existing models in the literature, e.g., Migon and Moura (2005), Migon and Penna (2006) and Amin and Salem (2015).

We start the development of our model considering the aggregate claims as in (4.9) and β_i as in (4.8). Next, we marginalize out β_i in the density of the aggregate claims, and hence we compress a two-level hierarchical structure into a single-level one. This procedure allows us to propose a mean parametrization over a distribution that holds the variation introduced by the rate hyperparameter β_i . Then, we can construct a hierarchical prior specification over the new parameters, incorporating parameter uncertainty and introducing partial pooling effects among the means as we showed in Section 3.2.3.

The resulting distribution, after marginalizing β_i (see Appendix A.1), is a 3-parameter distribution known as Generalized Beta Prime (GBP) (Dubey, 1970). Its probability density function is expressed as

$$f(s_i|\alpha^*, \gamma, \theta) = \frac{\Gamma(\alpha^* + \gamma)}{\Gamma(\alpha^*)\Gamma(\gamma)} \frac{1}{\theta} \frac{(s_i/\theta)^{\alpha^*-1}}{(1 + s_i/\theta)^{\alpha^*+\gamma}}, \quad (4.10)$$

where $\alpha^* = n_i\alpha_i$ and $\gamma > 0$ are shape parameters, and $\theta > 0$ is a scale parameter. Special cases of the GBP are the Pareto distribution (Type II), when $\alpha^* = 1$, and the Beta of the Second Kind, when $\theta = 1$. The GBP introduces flexibility that can accommodate the distribution for different medical services, an interesting feature for the application of this thesis. Furthermore, (4.10) could be directly used to model aggregate claims, however it is one of our goals to create a parametrization that is suitable for the borrowing strength property. We present next this parametrization.

4.4 Mean-Dispersion parametrization of the Generalized Beta Prime

This section introduces a mean-dispersion parametrization of the GBP, which is a cornerstone of the model developed in Chapter 5. In the literature, Ferrari and Cribari-Neto (2004) and Grün et al. (2011) presented a mean-based parametrization for the Beta distribution. However, to our knowledge, this has not yet been explored for the GBP.

Let X be a random variable GBP distributed, $X|\alpha, \gamma, \theta \sim GBP(\alpha, \gamma, \theta)$, with an expected value expressed as

$$E(X|\alpha, \gamma, \theta) = \frac{\theta\alpha}{\gamma - 1}, \quad \gamma > 1, \quad (4.11)$$

and, variance

$$V(X|\alpha, \gamma, \theta) = E(X|\alpha, \gamma, \theta)^2 \frac{\gamma + \alpha - 1}{\alpha(\gamma - 2)}, \quad \gamma > 2. \quad (4.12)$$

Then, from (4.12) we can define, similarly to the structure of Generalized Linear Models (Ohlsson and Johansson, 2010), a dispersion parameter as

$$\phi = \frac{\gamma + \alpha - 1}{\alpha(\gamma - 2)}, \quad \phi > 0. \quad (4.13)$$

It follows from (4.11) and (4.13), denoting the expected value as μ , that the original

parameters can be written as

$$\alpha = \frac{\gamma - 1}{\phi\gamma - 2\phi - 1}, \quad (4.14)$$

and

$$\theta = \mu(\phi\gamma - 2\phi - 1). \quad (4.15)$$

Under this parametrization, one can express X as follows

$$X|\mu, \gamma, \phi \sim GBP(\mu, \gamma, \phi), \quad \mu > 0, \gamma > 2, \phi > 0, \quad (4.16)$$

where μ is the individual claim amount mean, γ a shape parameter, and ϕ a dispersion parameter. Notice that $\gamma > 2$ guarantees the existence of the mean and dispersion.

Next section presents the individual risk model for dependent GBP distributed claim amounts. Once specified, we propose a hierarchical prior specification over the parameters introduced here.

4.5 Individual Risk Model under dependence

The individual risk model is defined as $S = X_1 + \dots + X_n$, where the number of individual components n is a known quantity, and in this case, each component represents a claim amount. This model can be used to specify the conditional aggregate claims as in (4.9).

The GBP random variable defined in Section 4.3 was obtained with a Compound Gamma-Gamma. Since the mixing distribution is on the rate parameter, one can apply the property of scale (rate) parameters (Klugman et al., 2012) to express the GBP as a ratio of Gamma distributed random variables. Therefore, the aggregate claims $S|\alpha^*, \gamma, \theta \sim GBP(\alpha^*, \gamma, \theta)$, whose density function is expressed as (4.10), can be written as

$$S = \theta \frac{G_{\alpha^*}}{G_{\gamma}}. \quad (4.17)$$

Given that $\alpha^* = n\alpha$, one can write the numerator as a sum of n identical and independent Gamma distributions with shape α

$$S = \theta \frac{G_{\alpha}^{(1)}}{G_{\gamma}} + \dots + \theta \frac{G_{\alpha}^{(n)}}{G_{\gamma}}, \quad (4.18)$$

where $G_{\alpha}^{(j)}|\alpha \sim \text{Gamma}(\alpha, 1)$, for $j = 1, \dots, n$, and $G_{\gamma}|\gamma \sim \text{Gamma}(\gamma, 1)$ are in-

dependent random variables with mean equal to α and γ , respectively. Notice that the individual claim amounts are now GBP distributed $X_j|\alpha, \gamma, \theta \sim GBP(\alpha, \gamma, \theta)$, for $j = 1, \dots, n$. The common random variable $G_\gamma|\gamma$ in the denominator of each individual component introduces dependence in the model.

Sarabia et al. (2016) presented the individual risk model for dependent Pareto distributed individual components, and subsequently derived the distribution of the sum. Since the Pareto is a special case of the GBP (when $\alpha = 1$), the individual risk model (4.17) is also a generalization of the Pareto case. As we argued in Section 4.3, the GBP assumption for the individual components can accommodate a broader spectrum of insurance claim types.

The parametrization shown in Section 4.4 can be implemented in the individual risk model presented here. In order to write the original parameters as a function of the new parameters, following the data structure presented in Section 4.1, it is required that, for each S_i , there is a θ_i associated. Recalling (4.8), where the pulling effect is introduced by both hyperparameters γ and θ , now this effect is performed only by γ . However, with our parametrization we can introduce parameter uncertainty through the group-specific means. Furthermore, a partial-pooling effect over the means is created by an overall expected cost random variable, a quantity whose posterior distribution collapses the information from all the groups. Then, we can express the aggregate claims as follows

$$S_i|n_i, \alpha_i, \gamma, \theta_i \sim GBP(n_i\alpha_i, \gamma, \theta_i),$$

where

$$n_i\alpha_i = \frac{\gamma - 1}{\phi_i\gamma - 2\phi_i - 1},$$

and

$$\theta_i = \mu_i(\phi_i\gamma - 2\phi_i - 1).$$

Then, we build a hierarchical prior specification on mean parameters

$$\mu_i|\mu, \sigma_\mu^2 \sim N(\mu, \sigma_\mu^2),$$

where μ and σ_μ are hyperparameters of the μ_i 's. Figure 4.2 illustrates the relationship between parameters under this model structure. This model is a two-level hierarchical model where individual claim amounts, X_{ij} 's, are now GBP distributed. Additionally, one could also create a hierarchical prior specification over the dispersion parameters.

Figure 4.3 shows samples from the marginals and joint bivariate posterior distributions of the original parameters in the model developed in Chapter 5, which was

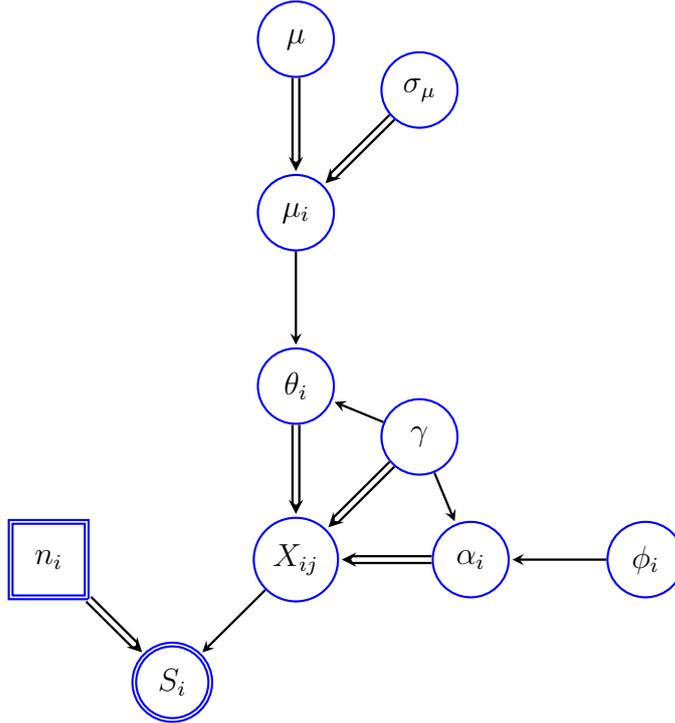


Figure 4.2: DAG Individual Risk Model with dependent GBP distributed claim amounts. The nodes represent quantities of interest. Square symbols represent known quantities and circles represent stochastic quantities. Single-arrows describe functional relationship, while double-arrows denote stochastic dependence. Double contour lines indicate observable quantities.

inspired by the development presented here. On the other hand, Figure 4.4 illustrates samples from the posterior distributions of the model parameters for the new parameters, μ , γ and ϕ . It is noteworthy that the latter parametrization avoids the negative correlation observed in Figure 4.3. Lastly, the resulting sparse joint parameter space facilitates the stochastic exploration of the sampling algorithm.

In this chapter, we studied aggregate loss models from a Bayesian standpoint. We explored the traditional Compound Poisson model and showed the development that was done by Migon and Moura (2005) on hierarchical prior specification, which was interestingly performed in the context of a health care plan. Then, we further developed the conditional aggregate claims leading to a GBP individual risk model under dependence, which better adapts to the ACA's context. This model is a generalization of models previously studied in Sarabia et al. (2016). Furthermore, we showed how to build a meaningful hierarchical structure over mean parameters, supported by a novel mean-dispersion parametrization of the GBP. In the next chapter, we apply the concepts studied here for the assessment of collective risks under the ACA.

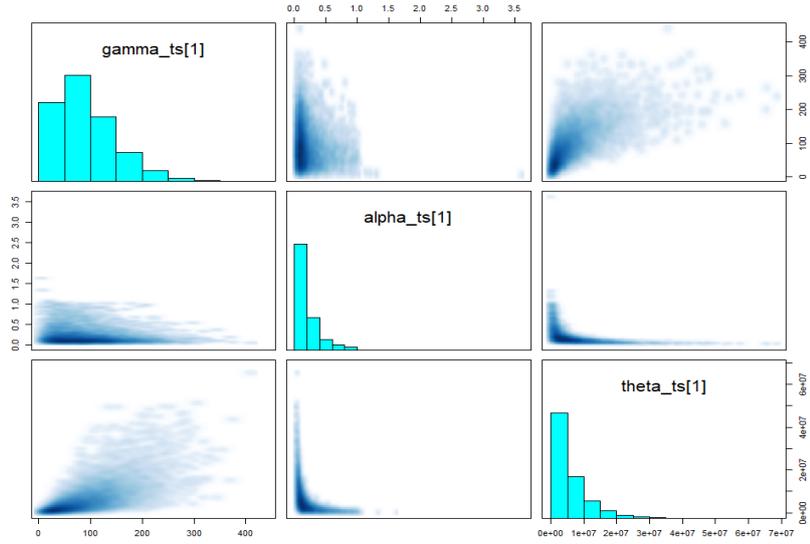


Figure 4.3: Posterior samples of γ , α and θ for *Injury and Poisoning* admissions in the Small Group segment, Arizona (AZ). This figure is a matrix, where the main diagonal represents the samples from the marginals and the other elements are samples from the joint bivariate between the parameter on the column and the row.

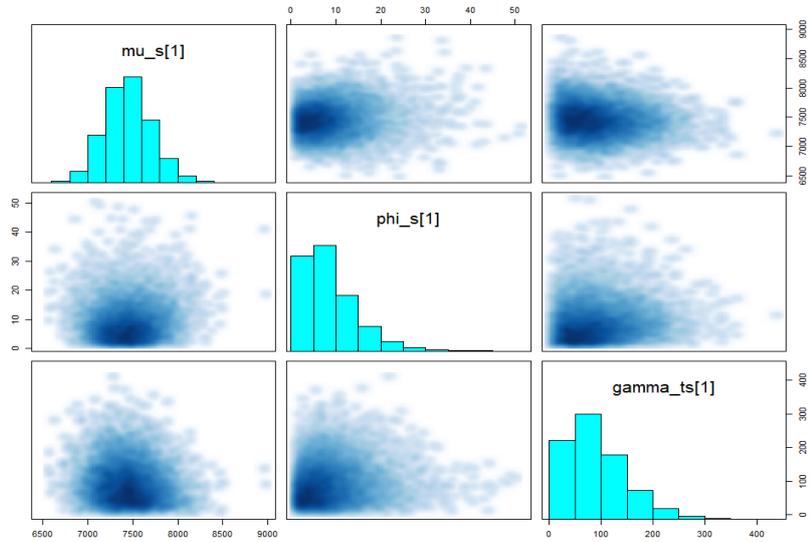


Figure 4.4: Posterior samples of γ , μ and ϕ for *Injury and Poisoning* admissions in the Small Group segment, Arizona (AZ). This figure is a matrix, where the main diagonal represents the samples from the marginals and the other elements are samples from the joint bivariate between the parameter on the column and the row.

Chapter 5

Collective risk assessment in ACA markets

In a health insurance contract, the insurer accepts the individual's actuarial risk related to health care utilization in exchange of a monthly premium. This risk is associated to demographic factors such as gender and age, existing health conditions, medical history and so on. Naturally, insurers would tend to use these factors to obtain fair actuarial premiums, i.e., insurance rates driven by the individual's expected health care cost. However, in practice, insurance regulations have restricted the use of risk factors in order to guarantee affordability and universal access to health insurance coverage. Thus, in these community rating schemes with unexisting or limited premium variation, governments implicitly impose cross subsidies from low-risk individuals towards the high-risk ones (Pupp, 1981).

In the US, the Affordable Care Act health care reform introduced major changes to the health insurance market. Since 2014, insurers in individual and small group markets are no longer allowed to reject customers (or impose extensive waiting periods in small group insurance), and premiums must be now determined according to a modified community rating scheme. The rating factors that can be used for pricing are age, smoking status (both with limits), geographical area, and family size. These new rules facilitated access to health insurance coverage for high risk individuals, especially those with pre-existing medical conditions, e.g., one can observe a significant decrease in the uninsured population by almost 10 million individuals in 2014 (KFF, 2019).

From the insurer's perspective, covering pre-existing medical conditions imply extra risk that is not charged individually. Generally, this would induce the insurer to increase the premium level for all its policyholders, leading to adverse selection and arbitrage opportunities in the market (Neuhaus, 1995). In order to avoid financial instability generated by high-risk individuals, the federal government has created a risk equalization program called Risk Adjustment (Kautter et al., 2014). Next, we explore the main characteristics of such program.

5.1 ACA Risk Adjustment

The ACA Risk Adjustment is a budget-neutral methodology used to normalize expected outcomes from different risk pools. It consists essentially of two steps: first, a carrier calculates a risk score for the enrollee population using a concurrent model proposed by the Health and Human Services (HHS) called HHS Hierarchical Conditions Categories (HHS-HCC). Then, the relative score, which represents the actuarial risk of the pool, is normalized with the market average actuarial risk. Given this normalization method, there are payer and receiver positions in the market (budget-neutral). Moreover, the risk that can be charged into premiums, using only age, geography, family size and smoking status, is also normalized and subtracted from the group risk score. Finally, the second step is to obtain absolute transfer amounts, which is achieved by scaling the final risk scores with statewide average premiums.

A simplified version of the transfer formula for a health care plan i is denoted as

$$T_i = \left(\frac{PLRS_i}{\sum_{i=1}^n s_i \times PLRS_i} - \frac{ARF_i}{\sum_{i=1}^n s_i \times ARF_i} \right) \times \bar{P}_s, \quad (5.1)$$

where T_i is the dollar amount that an insurer cedes/receives for plan i , $PLRS_i$ is the Plan Liability Risk Score, s_i is the enrollment market share relative to the statewide total enrollment, n the total number of insurers in the market pool, ARF_i (allowable rating factor) is the score that can be charged, and \bar{P}_s is the statewide average premium. In practice, the transfer formula considers other factors such as the geographical area, an induced demand factor and cost sharing, differences that Risk Adjustment tends to neutralize; the full version of the transfer formula can be found in Pope et al. (2014).

The subtraction in (5.1) represents the residual (relative) expected health care cost that is shared by all the insurers in the market, i.e., the risk that is spread to the statewide risk pool. This cost is associated to medical conditions that insurers cannot charge individually. The second term, \bar{P}_s , the weighted average premium of the statewide market, transforms the previously calculated relative scores into absolute transfer values. Then, a negative T_i results in a transfer to other insurers in the state, while a positive value is translated into a subsidy that the insurer receives from the others. Since this formula is normalized by all the risk pools within the statewide enrollment, transfer amounts sum to zero.

The normalizing PLRS in the denominator of (5.1) reflects the health status of the statewide population participating in the market. This score defines payer and receiver positions among insurers. For insurers on payer positions, an underestimation

of this risk score accompanied by low overall premiums, may lead to unexpected high payments to the market pool. The insurance unit Northwell Health CareConnect, participant of the small group market in New York, paid 11M to the market pool in 2015 and 112M in 2016 (Livingston, 2017). This generated high financial losses to the company, driving it to a shutdown.

For insurers participating in individual and small group markets, there are three sources that can generate a significant financial impact on revenues. The first one is related to the capacity of the insurer in identifying diagnoses on its own enrollee population, that is, calculating a PLRS that reflects its own population risks. The second source is the accuracy of the HHS-HCC model in risk scoring. There have been several discussions on how to improve this model; see, e.g., Centers for Medicare and Medicaid Services (2019). The third source, on which we focus in this thesis, is the collective risk associated to the statewide enrollment.

The assessment of statewide collective risks can ultimately help to understand, first, the behavior of the market pool PLRS and second, the statewide average premium, \bar{P}_s in (5.1). The market PLRS is affected by the enrollee population, which can suffer significant changes from year to year, and the average premium is linked to cost of medical services in the state. One should expect a great variation of PLRS among different market pools, given the existing differences in enrolment (risk exposure), as well as of average premiums, due to different cost structures in the states.

In the following section, we show how the Bayesian approach is particularly suited for this problem because of its ability to explicitly capture different levels of variation, in addition to handle some particularities of statewide data, e.g., missing or partial information. Lastly, all the variability is reflected over per-member-per-month cost estimates, which summarizes the health care cost allocation in each state. Moreover, full posterior distributions can indeed provide information about the stability of the statewide market and anticipate adverse outcomes in health care spending at state level.

5.2 Commercial Health Care Data Extract 2009-2015

The analysis presented in this chapter is made on the Commercial Health Care data extract published by the Society of Actuaries (2019) in collaboration with the Health Care Cost Institute. This data contains information related to health care cost and utilization in US health insurance markets during the period 2009 to 2015.

Table 5.1 shows 5 rows of a reduced version of the dataset. The column pool ID is an identifier for a combination of market segment, state and gender. Unfortunately, the factor age was not provided. The columns Membermonths, Admits and Cost represent the exposure, number of inpatient admissions and aggregate claims, respectively. In the following sections, these are denoted as $(m_{tmsgj}, n_{tmsgj}, s_{tmsgj})$ where t is the year, m market segment, s state, g gender and j service subcategory 2.

Furthermore, the Commercial Health Care (CHC) data extract reaches a granular level of medical services subcategories, in three different market segments, for twenty-one states and seven consecutive years, generating a total of 111.380 rows. However, in a few states with low exposure, especially in individual markets, no outcomes were provided; see Natsis (2019). The adopted model in Section 5.2.2 is calibrated for inpatient services at their last subcategory level, which contains more detailed information about claim types.

Table 5.1: Sample of the Commercial Health Care dataset

Year	Pool ID	Service	Subcat. 1	Subcat. 2	Member months	Admits	Cost
2009	1	Inpatient	Surgical	Circulatory	1.205.772	61	2.246.079
2009	1	Inpatient	Surgical	Digestive	1.205.772	200	2.936.328
2009	1	Inpatient	Surgical	Injury & Poisoning	1.205.772	121	2.617.554
2009	2	Inpatient	Surgical	Circulatory	2.933.580	471	17.289.373
2009	2	Inpatient	Surgical	Digestive	2.933.580	912	14.209.212

5.2.1 Exploratory data analysis

The CHC dataset is mainly divided in three market segments that operate in the US health system: large group, small group and the individual segment. The first two are advocated to employer-sponsored insurance, while the latter contains private insurance offered to individuals or families. The ACA Risk Adjustment, presented in the previous chapter, is implemented in individual and small group markets, working independently in each of the states. Therefore, our analysis is centered in these two segments.

The main variables of the dataset are the following:

- Year: 2009, 2010, 2011, 2012, 2013, 2014 and 2015
- Market segment: Large group, small group and individual
- Gender: Male, Female and All (unidentified gender)

- State: Arizona, California, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Maryland, Michigan, Minnesota, Missouri, Nevada, New York, Ohio, Oklahoma, Pennsylvania, Texas, Utah, Virginia, Wisconsin.
- Service category: Inpatient, Outpatient, Pharmaceutical and Professional
- Service subcategory 1: First level of classification depending on service category
- Service subcategory 2: Second level of classification depending on service subcategory 1.

The number of medical services in subcategories one and two is extensive; a full description can be found in Natsis (2019). An actuarial indicator to measure the allocation of health care resources in a state is Allowed PMPM (denoted as P), defined as the gross claim cost per-member-per-month¹. Considering the aforementioned factors, it can be written as

$$P_{tmsgj} = \frac{S_{tmsgj}}{m_{tmsg}}, \quad (5.2)$$

where t, m, s, g, j are the year, market, state, gender and service subcategory 2, respectively. The numerator $S_{tmsgj} = \sum_{i=1}^{N_{tmsgj}} X_{itmsgj}$ represents the aggregate claims, with X_{itmsgj} the i^{th} gross incurred claim amount and N_{tmsgj} the number of medical events, and m_{tmsg} the exposure measured by member-months. Thus, P_{tmsgj} represents the stochastic claim dollar amount per member per month spent on medical service j by risk pool $tmsg$. It should be mentioned that group $tmsg$ is exposed to the utilization of any medical service category, i.e., the exposure of a given risk pool is the same for all the services.

Furthermore, Allowed PMPM can be expressed as the product of frequency and severity, which can be obtained by including the number of claims N_{tmsgj} in (5.2),

$$P_{tmsgj} = \frac{S_{tmsgj}}{N_{tmsgj}} \times \frac{N_{tmsgj}}{m_{tmsg}}, \quad (5.3)$$

then it follows that

$$P_{tmsgj} = CPS_{tmsgj} \times fr_{tmsgj}, \quad (5.4)$$

where CPS_{tmsgj} is the average cost per claim and fr_{tmsgj} the frequency of claims for risk pool $tmsg$ on medical service j .

¹In this thesis, claim and medical event (e.g., inpatient admission) are used interchangeably, although, these terms might have different meanings in other contexts.

Figure 5.1 shows the evolution of P_m from 2009 to 2015, that were obtained from $\{s_{2009,m}/m_{2009,m}, \dots, s_{2015,m}/m_{2015,m}\}$, thus collapsing the cost of every health care service and risk pools. In the year 2014, as argued in Section 5.1, individual markets suffered significant changes in benefits and risk compositions. These changes provoked a massive increase of per-member-per-month costs, as observed in the figure. Figure 5.2 expands this analysis to the first category level of medical services. The trend in the individual market segment is explosive in every category during 2014 and 2015, with a slightly higher increase of inpatient services in 2014 compared with the others. At this point, however, it is not possible to address questions about within-service variation, which might be generated at a very low service level (e.g., increasing number of hospitalizations due to circulatory conditions).

Generally, *Inpatient*, *Outpatient*, *Pharmaceutical* or *Professional* services are significantly different from each other. This requires the modeling framework to be calibrated for each category. The model proposed in this thesis is applied to the *Surgical and Transplant* subcategory within *inpatient* services. Figure 5.3 shows the breakout at service subcategory two. Insurance costs related to a hospital admission due to a surgery or transplant are usually higher than other medical services. During the in-hospital stay, several health care resources are used, e.g., operating rooms, doctors' fees, hospital expenses, among others. Given the wide range of medical conditions and their severity, the total cost of a hospital admission may vary significantly from case to case.

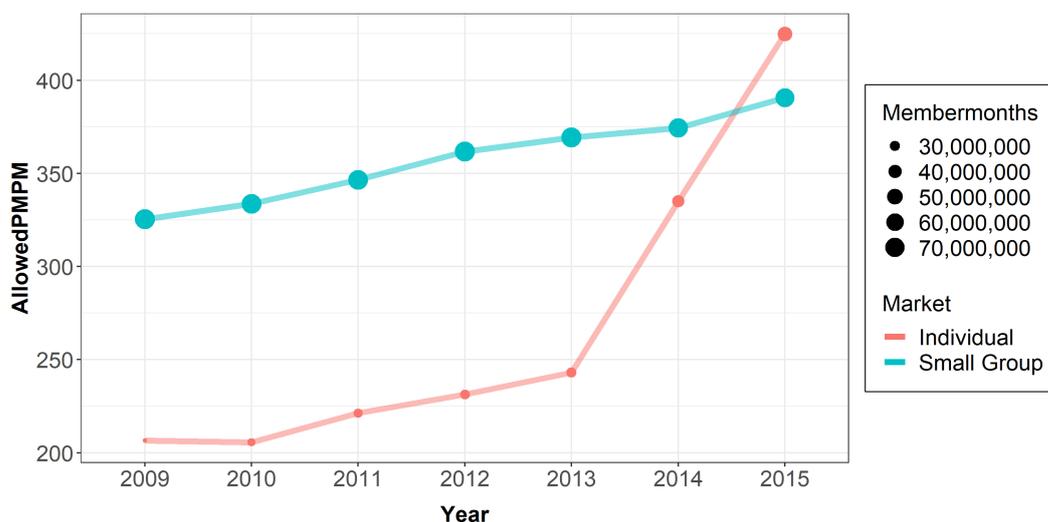


Figure 5.1: Evolution of Allowed PMPM (in USD) and member-months by market segment for the period 2009 - 2015

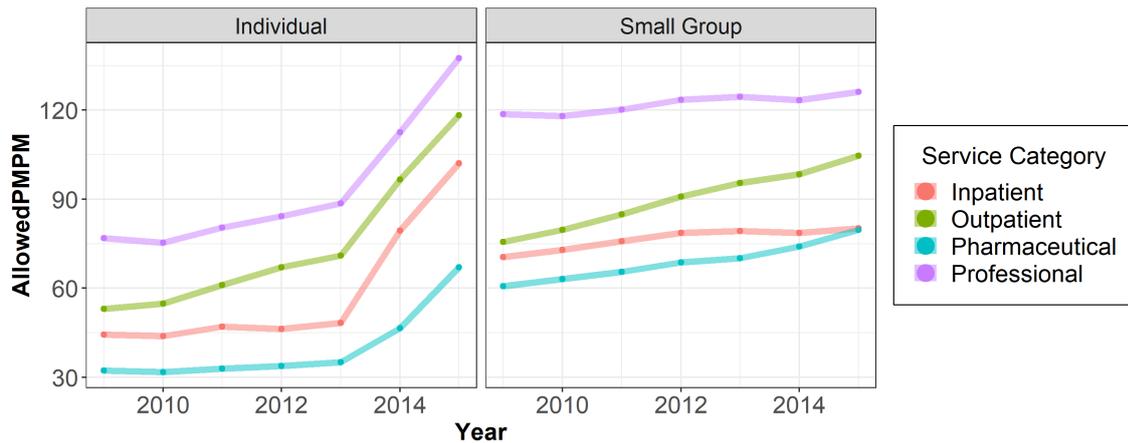


Figure 5.2: Evolution of Allowed PMPM (in USD) by service category for the period 2009 - 2015

Figure 5.4 illustrates the observed Allowed PMPM during the year 2015 in *Inpatient* services, the first subcategory level. The individual segment shows a higher cost allocation, especially for *Medical* and *Surgical and Transplant* services. This extra amount is stable across the second-level of service subcategories, as it can be seen in Figure 5.5. However, when looking at states separately, AllowedPMPM becomes highly volatile. Figure 5.6 shows this indicator for *Surgical and Transplant* services by states. Indiana (IN), Georgia (GA), Florida (FL), California (CA), Wisconsin (WI), Texas (TX) and New York (NY) show a higher allocated dollar amount in the individual segment compared with the small group. One reason of these differences is

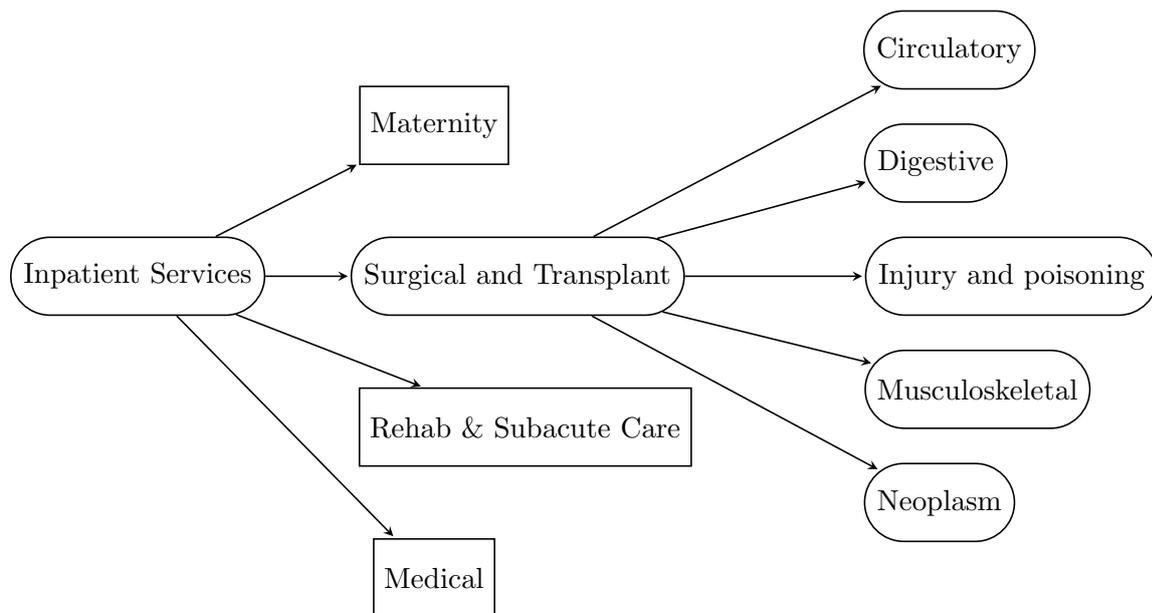


Figure 5.3: Inpatient - Surgical and Transplant breakout. Service category (1st layer), service subcategory 1 (2nd layer) and service subcategory 2 (3rd layer)

the significant exposure imbalance present in the CHC dataset; see tables B.1 and B.2 in Appendix B.1. Another driver is essentially the inclusion of previously uninsured population, which has followed different pace in each of the states. Moreover, the Medicaid expansion has contributed to increase this variation (KFF, 2019).

Particularly, the increasing number of enrollees with pre-existing conditions reshaped the demand of health care services. Since the implementation of a modified community rating, introduced in Section 5.1, the insurance cost incurred by this group has been spread to the statewide pool. This has been supported by the Risk Adjustment program, which has partially linked the performance of insurers to the stability

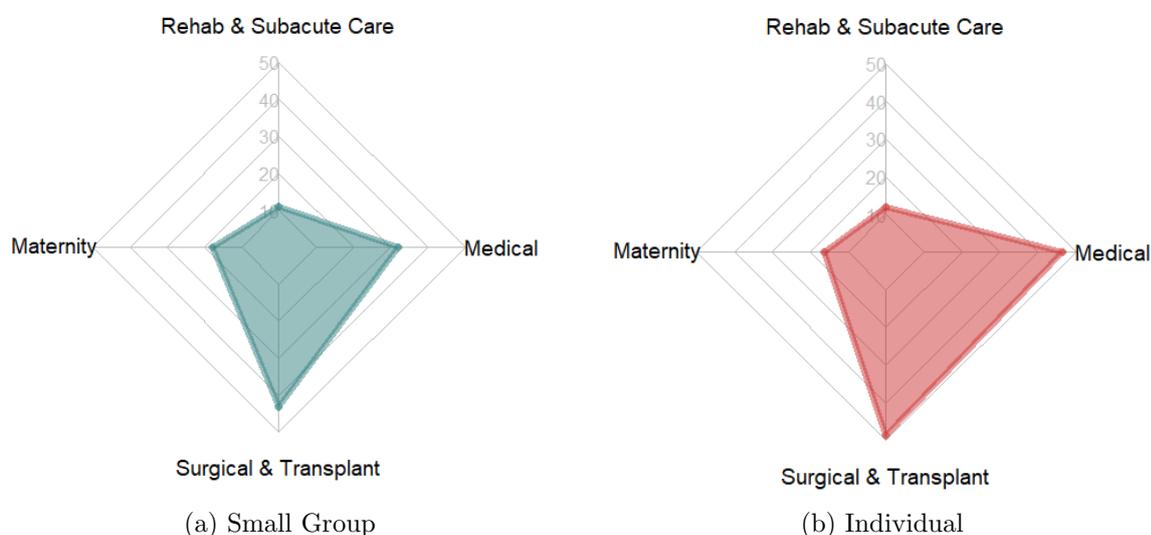


Figure 5.4: Allowed PMPM (in USD) for the first-level subcategory of *Inpatient* services by market segment in 2015

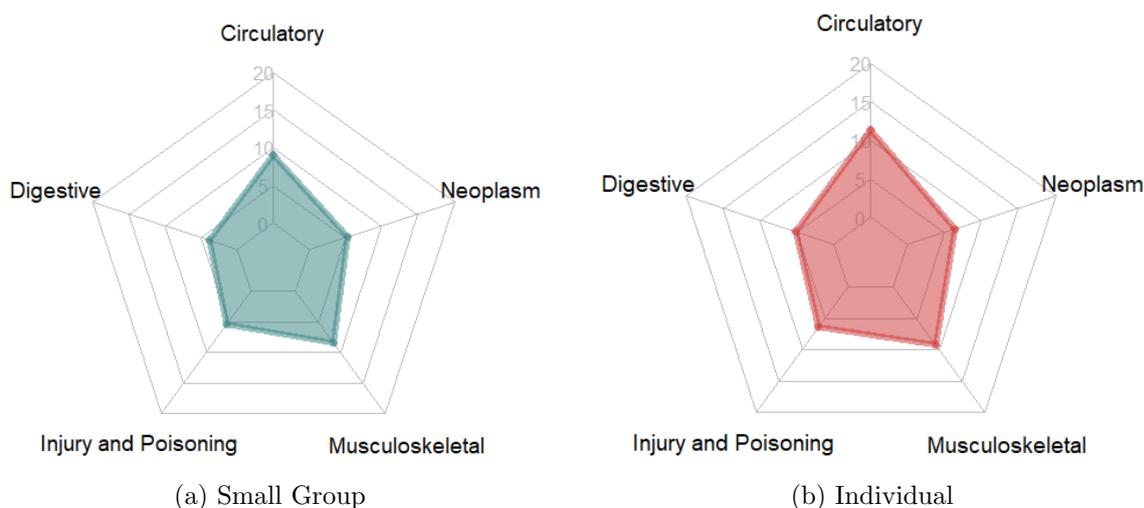


Figure 5.5: Allowed PMPM (in USD) for the second-level subcategory of *Surgical and Transplant* services by market segment in 2015

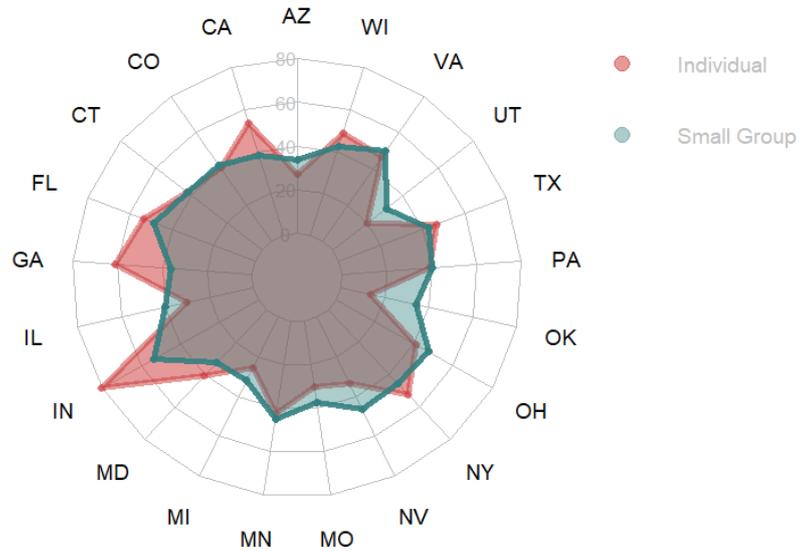


Figure 5.6: Allowed PMPM (in USD) for *Surgical and Transplant* in the 21 states by market segment in 2015

of the market as a whole. The shutdown of Northwell Health CareConnect plans in New York, with 44 percent of its revenues drained in Risk Adjustment payments (Livingston, 2017), and the closure of HealthyCT in Connecticut (Zorn, 2016) are just examples of market instability.

Recalling (5.4), we explore next how the average cost per claim and frequencies behaved in the period 2009-2015. Figure 5.7 shows the observed CPS_{tmj} , collapsing gender and states information. Furthermore, Figure 5.8 illustrates CPS_{tmsgj} , thus showing the outcomes of different risk pools *Male*, *Female* and *All* in a pre-selected group of states: Arizona (AZ), Colorado (CO), Florida (FL) and Texas (TX). In this graph, one can identify different levels of variation in the data: 1) among risk pools (*Male*, *Female* and *All*) within the same year, market segment, state and service; 2) across states within the same market segment and service; 3) along the years, within the same market segment, service and state.

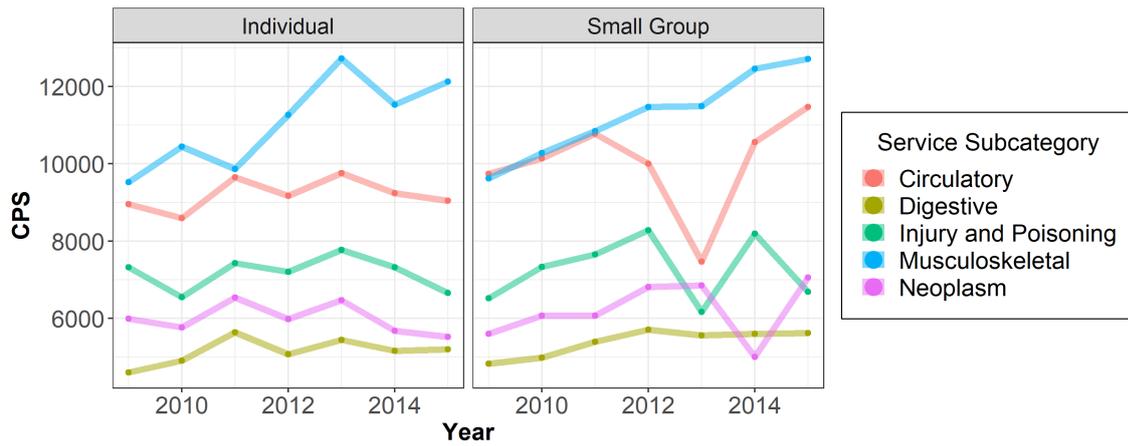


Figure 5.7: Evolution of the observed cost per admission (CPS) (in USD) by *Surgical and Transplant* service subcategory for the period 2009 - 2015

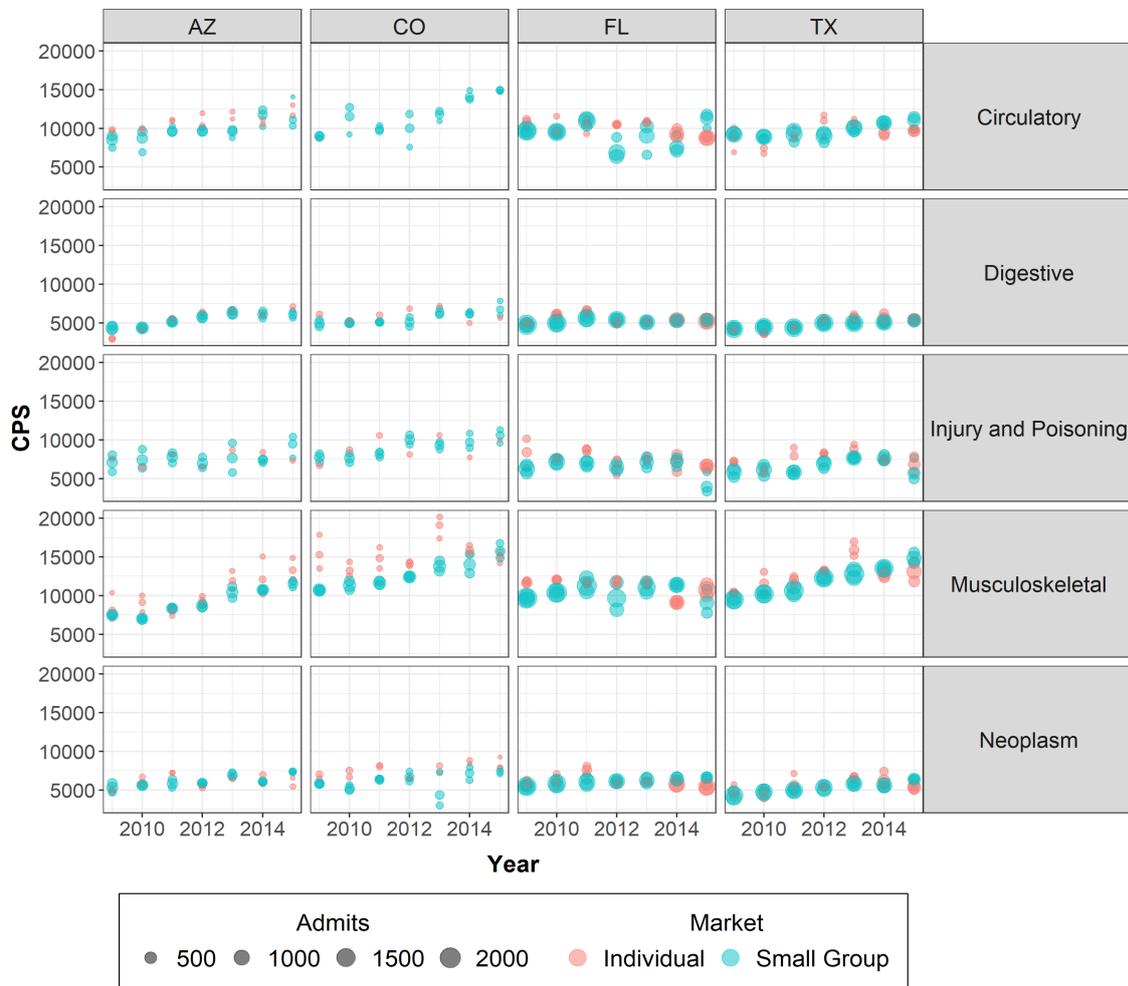


Figure 5.8: Cost per admission (CPS) (in USD) by *Surgical and Transplant* service subcategory in Arizona (AZ), Colorado (CO), Florida (FL) and Texas (TX) for the period 2009 - 2015. Each circle represents a risk pool for *Male, Female* or *All*.

The first point of variation, collectives within a statewide market in a year, helps to identify the behavior of the unobservable individual claim costs. A higher uncertainty among risk pools can be translated in a higher variation at individual claim level. Moreover, the outcomes of the pools are related to the statewide risk composition for that year, a feature that the adopted model aims to capture and a critical point for Risk Adjustment analysis. The second point of variation is explicitly model with a hierarchical parameter structure, as presented in Section 3.2.2. The last point of variation, interpreted by the evolution of the costs along the years, is critical to make out-of-sample analysis. This last point is not addressed in this thesis.

Similarly, Figure 5.9 and 5.10 shows the frequency of admissions per mille member-months in the individual and small group market segments for $t = 2009, \dots, 2015$, fr_{tmj} and fr_{tmsgj} (in thousands) respectively. Frequency of admissions is expected to be more volatile than *CPS* given the strong link with the health status of the population. The massive change of risk composition in the individual market in 2014 is fully reflected in the increasing utilization of medical services. However, this change was unstable across the states, as previously stated. For example, Texas, a state with high exposure, presented a slightly decrease in 2015 on *Injury and Poisoning* in the individual segment, compared with the increasing overall trend. Moreover, Florida did not show a drop on *Musculoskeletal*. In contrast, in the small group segment, the trends among states are similar. Generally, less representative states show more unstable trends along the years.

Although *Male*, *Female* and *All* are not identifiable in Figure 5.10, one can observe a high variation among these risk pools, which implies different patterns of utilization. In circulatory services, this between-pools variation is higher than in other subcategories for every pre-selected state. However, since the CHC data extract does not provide information for other important factors such as the age, it is not possible to identify useful risk patterns.

In this section we reviewed the principal changes in small group and individual market segments during the period 2009 - 2015. Furthermore, we discussed different sources of uncertainty and how significant changes may unequally impact statewide markets. As showed in Figure 5.8 and 5.10, several levels of variations can be identified before the information is collapsed on the indicator Allowed PMPM. A univariate analysis of this indicator is useful to visualize trends on health care cost allocation, as in Natsis (2019), however it does not provide a full perspective of the variation in different components of the aggregate loss (e.g., frequency and severity). The objective of this thesis is to quantify this variation and reflect it on statewide Allowed PMPM posterior distributions.

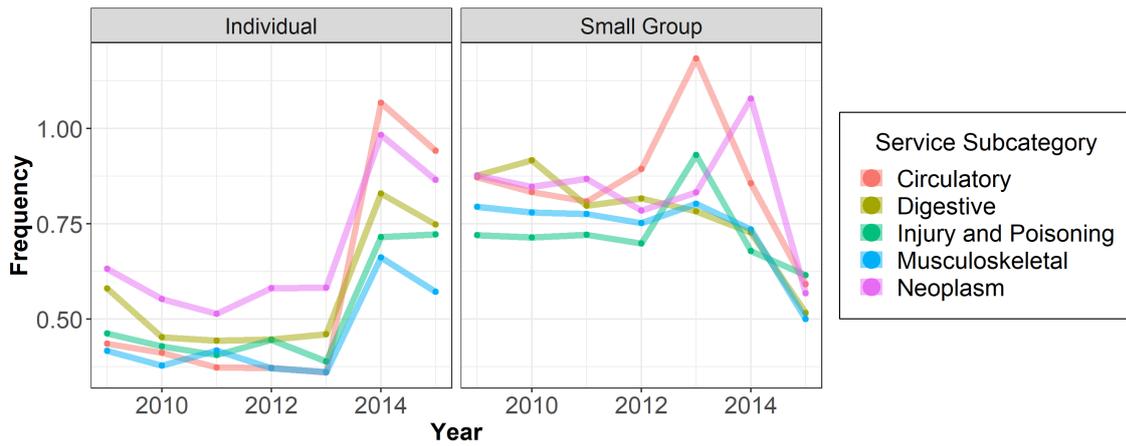


Figure 5.9: Evolution of the observed frequency per mille member-months by *Surgical and Transplant* service subcategory for the period 2009 - 2015

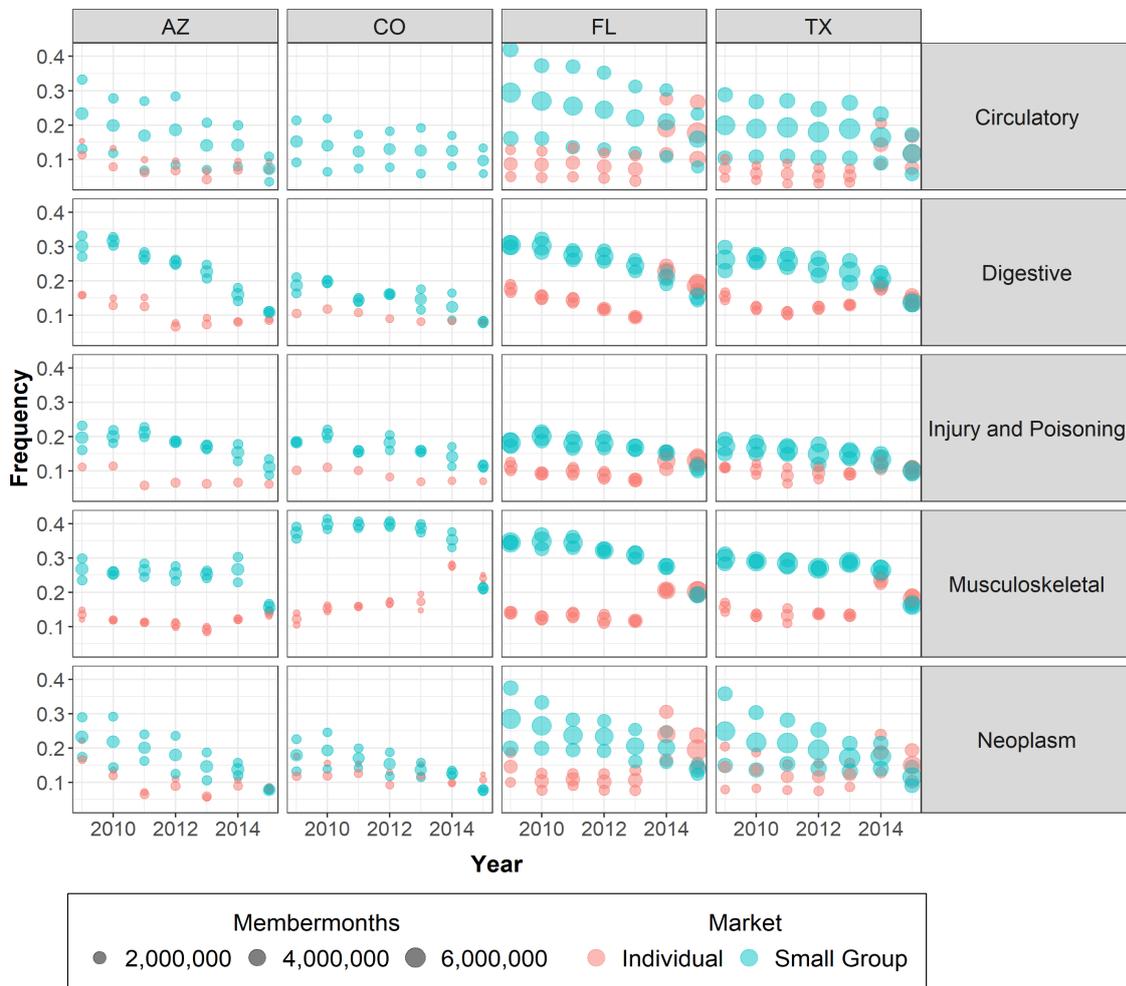


Figure 5.10: Inpatient frequency per mille member-months by *Surgical and Transplant* service subcategory in Arizona (AZ), Colorado (CO), Florida (FL) and Texas (TX) for the period 2009 - 2015. Each circle represents a risk pool for *Male, Female* or *All*.

5.2.2 Collective Risk Model: Compound Negative Binomial - Generalized Beta Prime

Let S_{tsg} be the aggregate claims for the risk pool identified with state s , gender g and year t . Market segment m and service subcategory j are fixed in the model, i.e., same framework is applied separately for each surgical and transplant service: *Circulatory, Digestive, Injury and poisoning, Musculoskeletal* and *Neoplasm*, in the individual and small group markets.

Therefore, the collective risk model can be expressed as

$$S_{tsg} = \sum_{i=1}^{N_{tsg}} X_{itsg}, \quad (5.5)$$

where tsg defines the risk pool, X_{itsg} is the i^{th} gross claim amount per hospital admission, and N_{tsg} the total number of admissions.

The model presented next is inspired by the development in Section 4.5. For the claims counts, we no longer assume a Poisson distribution but rather a Negative Binomial, similarly to Amin and Salem (2015). This is founded on the low number and high variation of hospital admissions at the second subcategory level. Our novelty approach is the development of the GBP claim amount distribution, presented in Section 4.3, and the mean-dispersion parametrization that was explored in Section 4.4. The model is then specified as follows

$$N_{tsg} | \delta_s, \beta_{tsg} \sim NB(\delta_s, \beta_{tsg}), \quad (5.6)$$

$$S_{tsg} | N_{tsg} = n_{tsg}, \alpha_{tsg}^*, \gamma_{ts}, \theta_{ts} \sim GBP(\alpha_{tsg}^*, \gamma_{ts}, \theta_{ts}), \quad (5.7)$$

where δ_s and β_{tsg} represent the shape and rate parameters of the Negative Binomial, respectively. The other parameters $\alpha_{tsg}^* = n_{tsg}\alpha_s$ and γ_{ts} are shape parameters, and θ_{ts} is the scale parameter of the GBP. As shown in Section 4.3 and 4.5, γ_{ts} and θ_{ts} are the shape and rate parameters of the Gamma hyperprior assumed for the rate parameter, and that was marginalized to obtain the GBP distribution. These common hyperparameters, within each state and year, aim to capture the dependence between all individual claims arising from the yearly statewide risk pool.

A convenient parametrization for the rate parameter in (5.6) (Ohlsson and Johansson, 2010) is expressed as follows

$$\beta_{tsg} = \frac{\delta_s}{m_{tsg}\zeta_s}, \quad \zeta_s > 0, \quad (5.8)$$

where m_{tsg} is the exposure based on member-months and ζ_s is the expected frequency of admissions in state s . Under this parametrization, a hierarchical structure is built over ζ_s ,

$$\zeta_s | \zeta, \sigma_\zeta^2 \sim \text{Normal}(\zeta, \sigma_\zeta^2), \quad (5.9)$$

where ζ represents the nationwide expected frequency in a given market and medical service, and σ_ζ^2 is the variance among state-specific frequencies. This prior specification introduces partial-pooling effects for statewide expected frequencies towards the nationwide expectation, which improves the estimates where the information is scarce (Gelman et al., 2007).

For the aggregate claims component S_{tsg} , we apply the mean-dispersion parametrization of the GBP introduced in Section 4.5. Therefore,

$$\alpha_{tsg} n_{tsg} = \frac{\gamma_{ts} - 1}{\phi_s \gamma_{ts} - 2\phi_s - 1}, \quad (5.10)$$

and

$$\theta_{ts} = \mu_s (\phi_s \gamma_{ts} - 2\phi_s - 1). \quad (5.11)$$

As presented in Section 4.5, we build a hierarchical prior specification for state-specific means

$$\mu_s | \mu, \sigma_\mu^2 \sim \text{Normal}(\mu, \sigma_\mu^2), \quad (5.12)$$

where μ represents the nationwide expected individual cost in a given market and medical service, and σ_μ^2 is the variance among state-specific average costs. This hierarchical structure introduces the borrowing strength property (explain in Section 3.2.3) among states. This information sharing is performed in each row of Figure 5.8.

Considering the CHC dataset and the observed values $\{m_{tsg}, n_{tsg}, s_{tsg}\}$ for $t = 2009, \dots, 2015$, the likelihood function is defined as

$$L(\delta_s, \beta_{tsg}, \alpha_{tsg}, \gamma_{ts}, \theta_{ts} | m_{tsg}, n_{tsg}, s_{tsg}) = \prod_{z=1}^Z f(s_{tsg[z]} | \alpha_{tsg}^*, \gamma_{ts}, \theta_{ts}) \times \quad (5.13)$$

$$\times p(n_{tsg[z]} | \delta_s, \beta_{tsg}, m_{tsg[z]}), \quad (5.14)$$

where $z = 1, \dots, Z$ is the observation identifier, β_{tsg} , $\alpha_{tsg}^* = \alpha_{tsg} n_{tsg}$ and θ_{ts} are the (transformed) original parameters, $f(s_{tsg[z]} | \alpha_{tsg}^*, \gamma_{ts}, \theta_{ts})$ the conditional density of the

aggregate claims

$$f(s_{tsg[z]}|\alpha_{tsg}^*, \gamma_{ts}, \theta_{ts}) = \frac{\Gamma(\alpha_{tsg}^* + \gamma_{ts})}{\Gamma(\alpha_{tsg}^*)\Gamma(\gamma_{ts})} \frac{1}{\theta_{ts}} \frac{(s_{tsg[z]}/\theta_{ts})^{\alpha_{tsg}^* - 1}}{(1 + s_{tsg[z]}/\theta_{ts})^{\alpha_{tsg}^* + \gamma_{ts}}}, \quad (5.15)$$

and $p(n_{tsg[z]}|\delta_s, \beta_{tsg}, m_{tsg[z]})$ the conditional mass probability function of the claim counts

$$p(n_{tsg[z]}|\delta_s, \beta_{tsg}, m_{tsg[z]}) = \frac{\Gamma(n_{tsg[z]} + \delta_s)}{\Gamma(n_{tsg[z]} + 1)\Gamma(\delta_s)} \left(\frac{\beta_{tsg}}{\beta_{tsg} + 1}\right)^{\delta_s} \left(\frac{1}{\beta_{tsg} + 1}\right)^{n_{tsg[z]}}, \quad (5.16)$$

where $m_{tsg[z]}$ is part of β_{tsg} through the transformation (5.8).

The log-likelihood function is then defined as

$$\begin{aligned} l(\delta_s, \beta_{tsg}, \alpha_{tsg}, \gamma_{ts}, \theta_{ts} | m_{tsg}, n_{tsg}, s_{tsg}) = & C + \sum_{z=1}^Z -\log(B(\alpha_{tsg}^*, \gamma_{ts})) - \log(\theta_{ts}) + \\ & + (\alpha_{tsg}^* - 1) \times \log(s_{tsg[z]}/\theta_{ts}) - (\alpha_{tsg}^* + \gamma_{ts}) \log(1 + s_{tsg[z]}/\theta_{ts}) + \log(\Gamma(n_{tsg[z]} + \delta_s)) - \\ & - \log(\Gamma(\delta_s)) + \delta_s \log\left(\frac{\beta_{tsg}}{\beta_{tsg} + 1}\right) + n_{tsg[z]} \log\left(\frac{1}{\beta_{tsg} + 1}\right), \quad (5.17) \end{aligned}$$

where C is a constant and

$$B(\alpha_{tsg}^*, \gamma_{ts}) = \frac{\Gamma(\alpha_{tsg}^*)\Gamma(\gamma_{ts})}{\Gamma(\alpha_{tsg}^* + \gamma_{ts})}$$

is the Beta function. Appendix C shows, in Stan language, a generic version of log-likelihood increments for a GBP model (see *functions* block).

Considering the functions of the transformed parameters (5.8), (5.10) and (5.11), the joint prior density of model parameters for state s and year t is expressed as

$$\begin{aligned} \pi(\delta_s, \zeta_s, \phi_s, \mu_s, \gamma_{ts}, \mu, \zeta, \sigma_\mu, \sigma_\zeta) = & \pi(\delta_s, \phi_s, \gamma_{ts}, \zeta_s, \mu_s | \zeta, \sigma_\zeta, \mu, \sigma_\mu) \times \\ & \times \pi(\zeta, \sigma_\zeta, \mu, \sigma_\mu), \quad (5.18) \end{aligned}$$

where $\pi(\delta_s, \phi_s, \gamma_{ts}, \zeta_s, \mu_s | \zeta, \sigma_\zeta, \mu, \sigma_\mu)$ is a first-level prior conditional on the hyperparameters $\zeta, \sigma_\zeta, \mu, \sigma_\mu$. Figure 5.11 illustrates the connections between observable quantities, first-level and second-level parameters. The node in the bottom represents the aggregate claims for risk pool tsg , which is generated by the frequency model on the left-hand side and the claim amount model on the right-hand side. The underlying exposure m_{tsg} is treated as a known quantity. This variable could be also modeled as in Migon and Penna (2006), however it is not in the scope of this thesis. The next section completes the prior specification with the fixed hyperparameters.

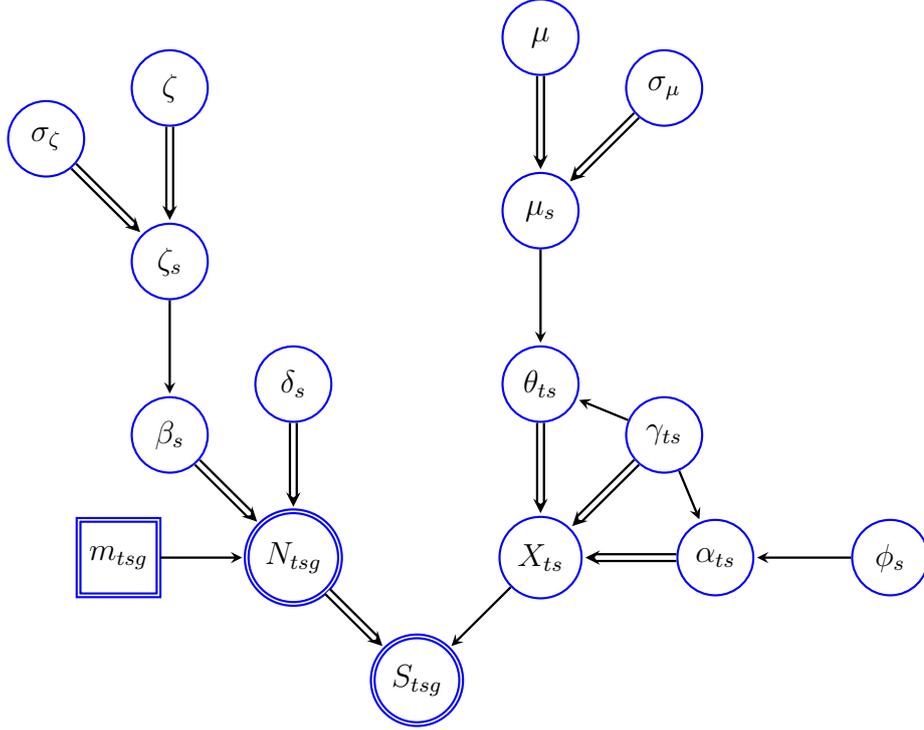


Figure 5.11: DAG Negative Binomial - Generalized Beta Prime Collective Risk Model with hierarchical prior structure. The nodes represent quantities of interest. Square symbols represent known quantities and circles represent stochastic quantities. Single-arrows describe functional relationship, while double-arrows denote stochastic dependence. Double contour lines indicate observable quantities.

5.2.3 Prior distributions

Given the parameter structure presented in (5.18), one must specify the corresponding hyperpriors. In this application, it is not our goal to inform model parameters with external data, although this could be possible. Then, we use weakly informative priors that do not have a significant impact on the posterior (Gelman et al., 2013). For hyperpriors on variance parameters, we follow the proposals made by Gelman et al. (2006).

Therefore, priors for the frequency model are defined as follows

$$\begin{aligned}
 \delta_s &\sim \text{Cauchy}^+(0, 1), \\
 \zeta_s | \zeta, \sigma_\zeta &\sim \text{Normal}^+(\zeta, \sigma_\zeta), \\
 \zeta &\sim \text{Normal}^+(\bar{F}r, \frac{s_{Fr}}{2}), \\
 \sigma_\zeta &\sim \text{Cauchy}^+(0, s_{Fr}),
 \end{aligned}$$

where $\bar{F}r$ and s_{Fr} are the observed overall frequency and standard deviation, weighted

by exposure.

Priors for the claim amount model are

$$\begin{aligned}\phi_s &\sim \text{Normal}^+(1, 25), \phi_s > 1, \\ \gamma_{ts} &\sim \text{Normal}^+(3, 100), \gamma_{ts} > 3, \\ \tilde{\mu}_s &\sim \text{Normal}(0, 3), \\ \mu &\sim \text{Normal}^+(C\bar{P}S, \frac{s_{CPS}}{2}), \\ \sigma_\mu &\sim \text{Cauchy}^+(0, s_{CPS}),\end{aligned}$$

where $C\bar{P}S$ and s_{CPS} are the observed overall average cost per claim and standard deviation, weighted by number of claims. A non-centered parametrization is implemented on the means, therefore $\mu_s = \mu + \sigma_\mu * \tilde{\mu}_s$. In the Stan code, Appendix C, one can find the prior specification in the block *model*.

Originally, the domain restrictions for the shape and dispersion parameters are $\gamma_{ts} > 2$ and $\phi_s > 0$, respectively. However, due to stability purposes in the HMC-NUTS sampling algorithm, we restrict these domains to $\gamma_{ts} > 3, \phi_s > 1$. As shown in Section 4.5,

$$\alpha_{ts}^* = \frac{\gamma_{ts} - 1}{\phi_s \gamma_{ts} - 2\phi_s - 1},$$

where $\alpha_{ts}^* > 0$, so it follows that

$$\phi_s > \frac{1}{(\gamma_{ts} - 2)}.$$

For the range of $2 < \gamma_{ts} < 3$ the sampling is unstable, due to the high values that the dispersion takes. The restriction of $\gamma_{ts} > 3$ does not significantly affect the inference since we expect γ_{ts} to be high.

It is worth mentioning that one could inform mean or variance hyperparameters, employing information such as the number of insurers in the market or the percentage of uninsured population, among others. For instance, an inclusion of a high proportion of previously uninsured population to the system can be incorporated on $P(\gamma_{ts})$, which leads to higher uncertainty on the aggregate claims and heavier tail. On the other hand, information about new expensive treatments that were not covered in past years, and that are offered nationwide, can be considered on $P(\mu)$ and $P(\zeta)$.

5.2.4 Posterior predictive sampling

The algorithm used to draw samples from the posterior distribution is the NUTS, introduced in Section 3.3. Once MCMC samples are obtained for all model parameters, we can draw values from the posterior predictive distributions. First, based on the exposure m_{tsg} , δ_s and ζ_s , one can obtain β_{tsg} using the transformation (5.8). Then draw from the number of admissions posterior predictive $N_{tsg}|\delta_s, \beta_{tsg}$, and sequentially, from the aggregate loss $S_{tsg}|N_{tsg} = n_{tsg}, \alpha_{ts}, \gamma_{ts}, \theta_{ts}$ (see Figure 5.11), using the transformations (5.10) and (5.11), and the ratio of Gamma distributions (4.17). Algorithm 1 summarizes this procedure, which we use to draw samples for every state. In the Stan code, Appendix C, one can find every parameter transformation in *Transformed parameters* and Algorithm 1 in *Generated Quantities*. The results are presented in the following section.

Algorithm 1 Drawing samples from the posterior predictive of the aggregate claims

```

1: input =  $I, m_{tsg}, \beta_{tsg}, \vec{\delta}_s, \vec{\theta}_{ts}, \vec{\gamma}_{ts}, \vec{\alpha}_{ts}$ 
2: for  $i = 1$  to  $I$  do
3:    $nrep[i] = \text{Generate } N_{tsg} \sim NB(\delta_s^{(i)}, \beta_{tsg}^{(i)})$ 
4:   if ( $nrep[i] == 0$ ) then
5:      $srep[i] = 0$ 
6:   else
7:      $g_{rng}^1 = \text{Generate } G_{\alpha_{ts}} \sim \text{Gamma}(\alpha_{ts}^{(i)} nrep[i], 1)$ 
8:      $g_{rng}^2 = \text{Generate } G_{\gamma_{ts}} \sim \text{Gamma}(\gamma_{ts}^{(i)}, 1)$ 
9:      $srep[i] = \theta_{ts}^{(i)} \frac{g_{rng}^1}{g_{rng}^2}$ 
10:  end if
11: end for

```

5.3 Numerical results

The model presented in the previous sections was fitted for *Circulatory*, *Digestive*, *Injury and poisoning*, *Musculoskeletal* and *Neoplasm*, for individual and small group markets. The NUTS sampling algorithm had no issues in the exploration of the posterior distribution for all of the services (e.g., no divergent transitions during the exploration). Chain trajectories have mixed for all model parameters; e.g., Figure B.1 in Appendix B.2 illustrates trace plots of hierarchical parameters for *Injury and poisoning* in the small group market segment. Figure 5.12 shows aggregate claims replications (i.e., draws from the posterior predictive distribution) in form of credible intervals for *Neoplasm*. Appendix B.2, Section B.2.1 shows the credible intervals for

the other services in the small group segment, while Section B.2.2 does it for the individual segment.

In Figure 5.12, one can observe that the model positively replicates the loss process in almost every state (i.e., observed values are not considerably far from the median), except for the year 2015 where the model overestimates the aggregate claims (e.g., AZ-Arizona). This location issue is mostly coming from the posterior predictive of the number of claims (see Figure B.3 in Appendix B.2). Furthermore, more volatile markets show wider credible intervals of the aggregate claims, e.g., in Figure 5.12, California (CA), Nevada (NV) and New York (NY), or in Figure B.9 and B.11 in Appendix B.2, the state of New York for 2015. The higher uncertainty has been passed through the parameter structure shown in Figure 5.11, and it has been generated by: the posterior predictive of the number of claims and unobservable quantities (e.g., nationwide parameters). This uncertainty can produce instability in the market and unexpected Risk Adjustment payments. As we mentioned in Section 5.1, Northwell Health CareConnect paid 11M to the market pool in 2015 and 112M in 2016 in the small group market of New York (Livingston, 2017).

Interestingly, we are able not only to pass the uncertainty from unobservable quantities in high hierarchical levels but also to make inferences about these quantities. These can provide relevant information about health care costs in the nationwide individual and small group market segments. Figure 5.13 shows posterior histograms of the samples drawn from the nationwide cost means, μ in the adopted model. The first insight is that every service in the individual market segment shows a higher uncertainty on expected mean costs than the small group (i.e., posteriors with heavier tails). *Circulatory* conditions are, on average, the most expensive in the *Surgical and Transplant* subcategory, and they also present, on average, the highest variation across state-specific means (see Figure B.2 in Appendix B.2).

We previously studied credible intervals of the aggregate claims posterior predictive. However, to make a comparison of allocated costs between states, it is necessary to show the results in a per-member-per-month basis. Figure 5.14 presents the posteriors of Allowed PMPM as expressed in (5.2), and the observed values in every service for Arizona (AZ), Colorado (CO), Florida (FL) and Texas (TX), for the year 2015. For most of the services and states, the distribution of per-member-per-month costs shows a greater skewness in individual markets. Generally, one could say that the high variation of inpatient services leads to posteriors with heavy tails in most of the statewide markets. The tail represents potential scenarios triggered from a high-risk yearly enrollment, resulting in a high number of expensive inpatient admissions.

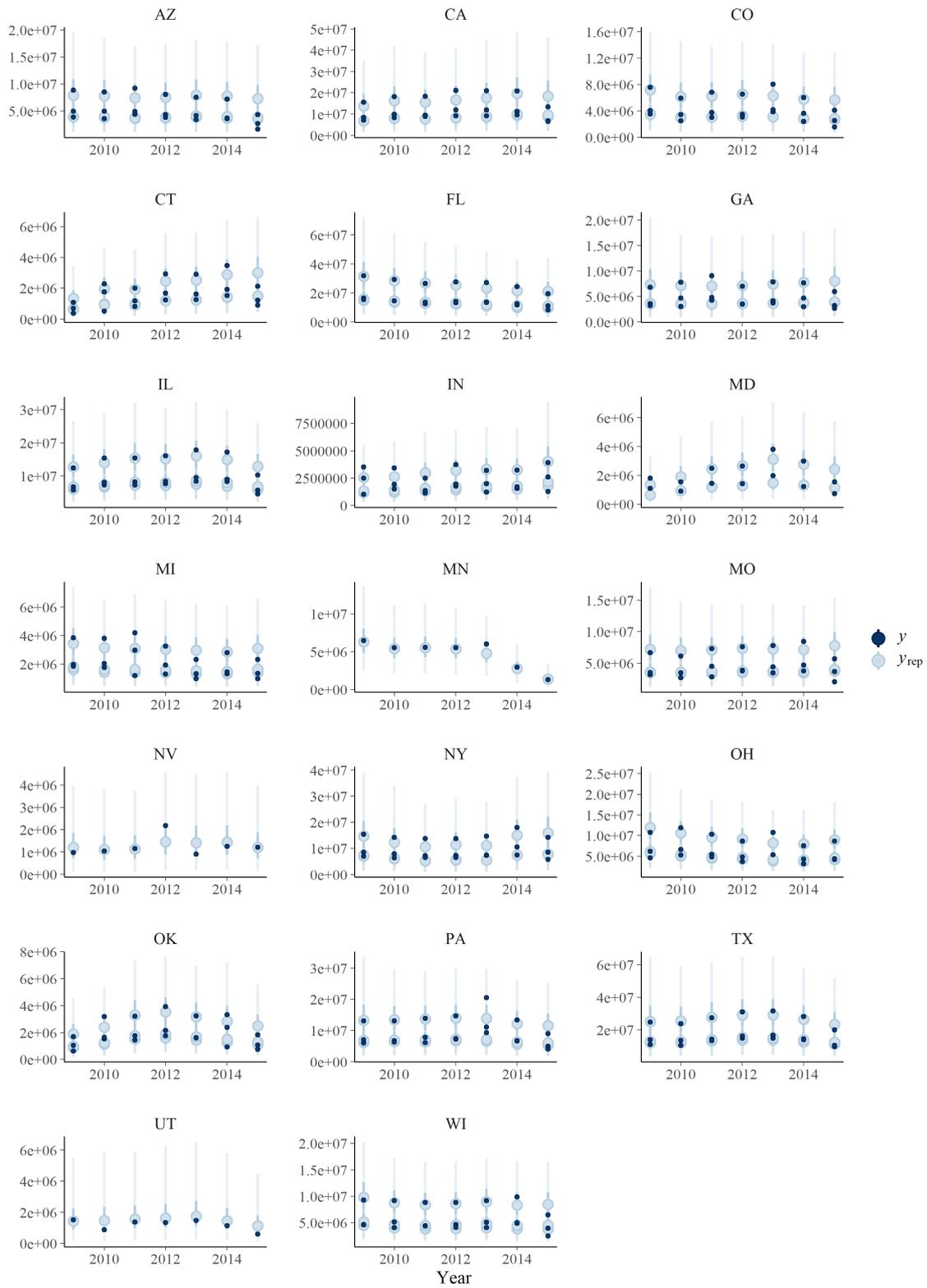


Figure 5.12: *Neoplasm* aggregate claims (in USD) 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

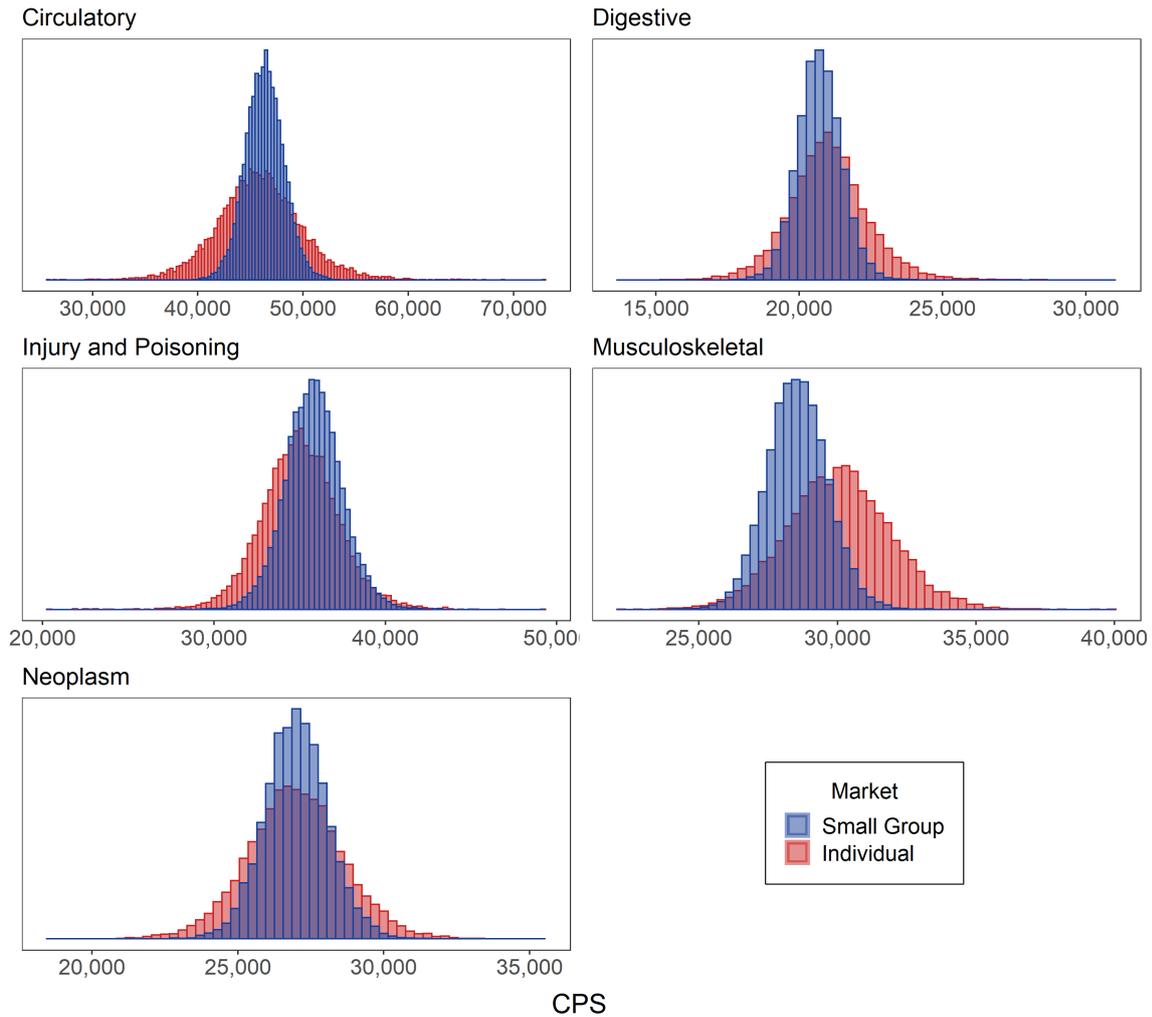


Figure 5.13: Posterior samples of μ , representing the Cost Per Service (CPS) (in USD), by service subcategory for the Small Group and Individual segments in the year 2015

In this chapter, we first reviewed the risk-sharing mechanism that the ACA Risk Adjustment introduced in individual and small group markets. Then, we motivated the assessment of statewide collective risks to anticipate the stability of the market and ultimately the insurer's relative risk position. In this context, setting appropriate rate levels is crucial to cover health risks from the state market pool. From their side, state markets have developed different experiences along the years, first, in terms of size (risk exposure), and second, in terms of frequencies and costs. In fact, we showed in this section that the uncertainty of inpatient cost allocation significantly differs from market to market (Figure 5.14). This strongly motivates the assessment of statewide collective risks, which are of great importance under the ACA dynamics.

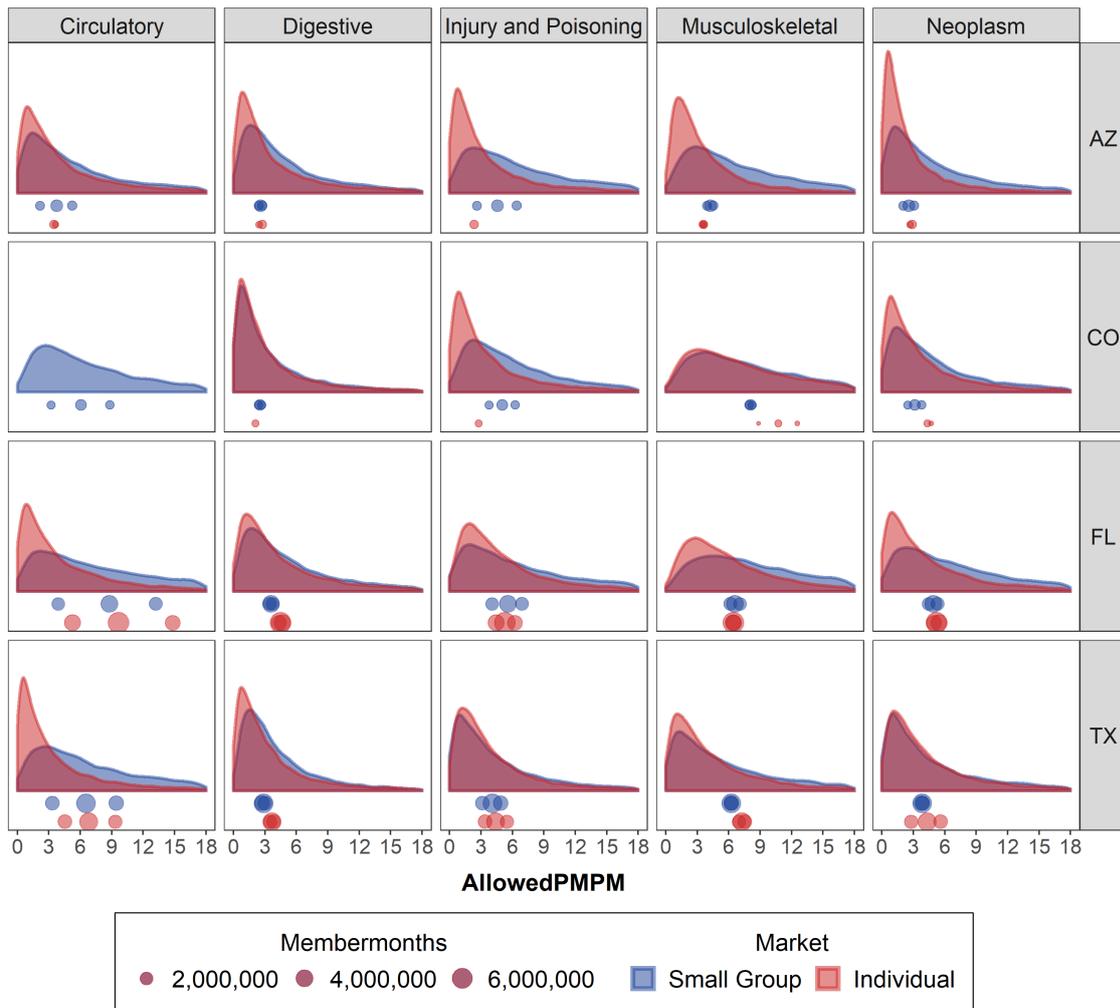


Figure 5.14: MCMC Posterior predictive densities of Allowed PMPM (in USD) of *Surgical and Transplant* in the small group and individual segments for 2015 (AZ-Arizona, CO-Colorado, FL-Florida and TX-Texas). Observed values for the same year are displayed below the densities.

Furthermore, we argued that the Bayesian approach is suited to this problem because of its ability to explicitly quantify uncertainty. It also provides a mechanism to inform parameters when there is important information not yet reflected in the data. This could be done by building an *a priori* dependence structure, as shown in the diagram 5.11, or by specifying fixed hyperparameters in the hyperpriors (Section 5.2.3). Next, we further describe some other thoughts and state our conclusions.

Chapter 6

Conclusions

In the first years after the ACA changes were implemented, the uncertainty of statewide collective risks provoked financial instability in individual and small group markets. In fact, a group of insurers suffered unexpectedly high losses due to adverse Risk Adjustment positions (Livingston, 2017, Zorn, 2016). In stable environments, the projection of statewide health care costs has been performed with univariate techniques (Natsis, 2019). However, the increasing uncertainty that the ACA incorporated has called on advanced modeling techniques.

The framework presented in this thesis aims to quantify important sources of uncertainty in the estimation of collective risks. We further develop previous works in the collective risk theory presented by Migon and Moura (2005), Migon and Penna (2006) and Amin and Salem (2015). We present a novelty distribution for the conditional aggregate claims: a 3-parameter distribution known as Generalized Beta Prime (GBP). This model is a generalization of the sum of dependent Pareto distributed claim amounts, presented by Sarabia et al. (2016). We present the GBP based on a mean-dispersion parametrization, which allows the introduction of a meaningful hierarchical prior specification. This actually gives flexibility to build prior dependencies on the mean and dispersion parameters.

Furthermore, we apply the model on a subset of the Commercial Health Care dataset published by the Society of Actuaries (2019), the *Surgical and Transplant* services breakout for the small group and individual market segments. The results presented here are: credible intervals for the replications of the aggregate claims; MCMC histograms of the nationwide cost means corresponding to *Circulatory, Digestive, Injury and poisoning, Musculoskeletal* and *Neoplasm*; full posteriors of the variation of state-specific means; and finally, posterior predictive distributions of the per-member-per-month costs in Arizona (AZ), Colorado (CO), Florida (FL) and Texas (TX).

One point for future development is the building of a full probabilistic model that reaches the statewide per-member-per-month claim cost, considering all medical

services and the subsequent aggregation. It is worth noting that the flexibility introduced by the GBP can accommodate the distribution shape for other service types. Additionally, the posteriors of per-member-per-month costs obtained from this model could be used to predict state average plan liability risk scores (PLRS), and hence help to anticipate relative Risk Adjustment positions.

Another point for future research is the link between the changes in the uninsured population or other external variables and the consequent new patterns in health care utilization. For this, it would be also necessary to account for more demographic factors such as the age, which is a critical driver of the population health status. Moreover, additional information with respect to the number of insurers and providers in statewide markets can be translated into more informative hierarchical priors in the severity component.

Following the first years of instability in individual and small group markets, the period from 2017 to 2019 showed signs of recovery and the situation of insurers improved substantially (Cox et al., 2019). However, in the present year, the COVID-19 pandemic is again increasing uncertainty in health insurance markets: individuals are shifting from small or large group markets towards individual markets; there are new trends in health care utilization and a deferred demand of medical services; and new treatments and comorbidities are expected to change the cost per medical service. Furthermore, a proportion of the costs associated to the pandemic are being waived with federal regulation (Centers for Medicare and Medicaid Services, 2020), putting pressure on the market pool.

Therefore, the changing patterns in health care utilization and the creation of risk-sharing rules will require a comprehensive assessment of statewide collective risks. Our findings can contribute to the development of risk models that take into account the new dynamics and emerging sources of uncertainty.

Appendix A

Generalized Beta Prime

A.1 GBP: Compound Gamma-Gamma

Let X follow a Gamma distribution, $X|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$, with shape-rate parametrization. Let β be a random variable following a Gamma $\beta|\gamma, \theta \sim \text{Gamma}(\gamma, \theta)$. Then, the unconditional density, obtained by integrating out β , is

$$\begin{aligned}
 f(x|\alpha, \gamma, \theta) &= \int_0^{+\infty} f(x|\alpha, \beta) f(\beta|\gamma, \theta) d\beta \\
 &= \int_0^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \frac{\theta^\gamma}{\Gamma(\gamma)} \beta^{\gamma-1} e^{-\theta\beta} d\beta \\
 &= \frac{x^{\alpha-1} \theta^\gamma}{\Gamma(\alpha) \Gamma(\gamma)} \int_0^{+\infty} \beta^\alpha e^{-\beta x} \beta^{\gamma-1} e^{-\theta\beta} d\beta \\
 &= \frac{x^{\alpha-1} \theta^\gamma}{\Gamma(\alpha) \Gamma(\gamma)} \int_0^{+\infty} \beta^{\alpha+\gamma-1} e^{-\beta(x+\theta)} d\beta \\
 &= \frac{x^{\alpha-1} \theta^\gamma}{\Gamma(\alpha) \Gamma(\gamma)} \frac{\Gamma(\alpha+\gamma)}{(x+\theta)^{\alpha+\gamma}} \int_0^{+\infty} \beta^{\alpha+\gamma-1} e^{-\beta(x+\theta)} \frac{(x+\theta)^{\alpha+\gamma}}{\Gamma(\alpha+\gamma)} d\beta \\
 &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha) \Gamma(\gamma)} \frac{x^{\alpha-1} \theta^\gamma}{(x+\theta)^{\alpha+\gamma}} \\
 &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha) \Gamma(\gamma)} \frac{x^{\alpha-1} \theta^\gamma \theta^\alpha}{(x+\theta)^{\alpha+\gamma} \theta^\alpha} \\
 &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha) \Gamma(\gamma)} \frac{x^{\alpha-1}}{\frac{(x+\theta)^{\alpha+\gamma} \theta^{\alpha+\gamma}}{\theta^{\alpha+\gamma}}} \\
 &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha) \Gamma(\gamma)} \frac{1}{\theta} \frac{(x/\theta)^{\alpha-1}}{(1+x/\theta)^{\alpha+\gamma}} \\
 &= \frac{1}{B(\alpha, \gamma)} \frac{1}{\theta} \frac{(x/\theta)^{\alpha-1}}{(1+x/\theta)^{\alpha+\gamma}}
 \end{aligned}$$

where $B(\alpha, \gamma)$ is the Beta function and $f(x|\alpha, \gamma, \theta)$ has the form of a Generalized Beta Prime distribution with two shape parameters α and γ , and one scale parameter θ .

Appendix B

CHC dataset

B.1 Actuarial indicators

Table B.1: Actuarial indicators for Surgical and Transplant in the small group segment by state in 2015

Market	State	Cost	Membermonths	AllowedPMPM
Small Group	AZ	112.784.125	3.344.264	33,72
Small Group	CA	239.986.579	6.244.710	38,43
Small Group	CO	109.776.958	2.622.546	41,86
Small Group	CT	52.402.366	1.242.452	42,18
Small Group	FL	383.349.882	7.950.252	48,22
Small Group	GA	142.876.775	3.965.168	36,03
Small Group	IL	191.786.060	4.758.020	40,31
Small Group	IN	78.611.725	1.464.040	53,70
Small Group	MD	36.490.499	1.128.004	32,35
Small Group	MI	41.609.034	1.296.812	32,09
Small Group	MN	26.161.217	578.670	45,21
Small Group	MO	141.826.801	3.797.928	37,34
Small Group	NV	28.073.998	601.494	46,67
Small Group	NY	265.944.268	5.821.508	45,68
Small Group	OH	165.006.602	3.473.108	47,51
Small Group	OK	37.400.485	1.096.248	34,12
Small Group	PA	169.143.111	4.195.618	40,31
Small Group	TX	441.195.684	10.350.922	42,62
Small Group	UT	24.991.474	820.588	30,46
Small Group	VA	32.350.093	649.330	49,82
Small Group	WI	154.078.862	3.601.648	42,78

Table B.2: Actuarial indicators for Surgical and Transplant in the individual segment by state in 2015

Market	State	Cost	Membermonths	AllowedPMPM
Individual	AZ	36.289.970	1.352.274	26,84
Individual	CA	23.143.104	429.744	53,85
Individual	CO	34.656.738	856.492	40,46
Individual	CT	14.342.149	335.020	42,81
Individual	FL	697.964.940	13.094.660	53,30
Individual	GA	483.176.256	7.924.618	60,97
Individual	IL	25.240.930	833.802	30,27
Individual	IN	87.089.660	1.083.470	80,38
Individual	MD	23.828.292	580.956	41,02
Individual	MI	29.270.790	1.153.222	25,38
Individual	MN	5.009.576	117.468	42,65
Individual	MO	49.444.802	1.619.952	30,52
Individual	NV	11.457.618	341.410	33,56
Individual	NY	54.814.035	1.044.810	52,46
Individual	OH	61.015.079	1.479.476	41,24
Individual	OK	2.908.368	213.862	13,60
Individual	PA	83.062.151	2.131.986	38,96
Individual	TX	424.891.295	9.140.948	46,48
Individual	UT	18.219.719	931.374	19,56
Individual	VA	21.556.163	458.498	47,01
Individual	WI	36.466.974	745.888	48,89

B.2 Model outputs

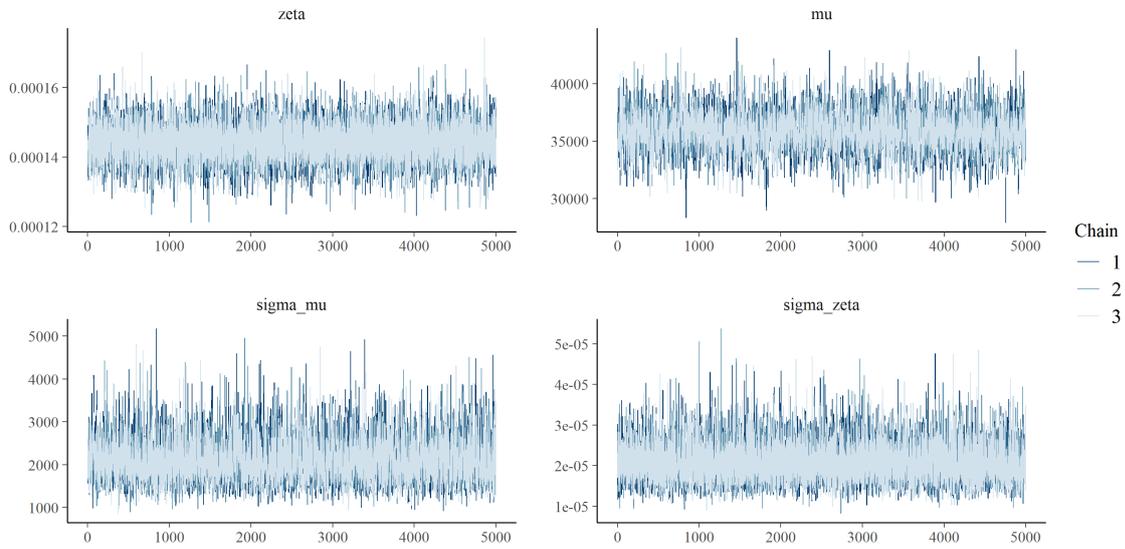


Figure B.1: Chains trajectories of ζ , μ , σ_μ and σ_ζ for the *Injury and Poisoning* model in the small group segment

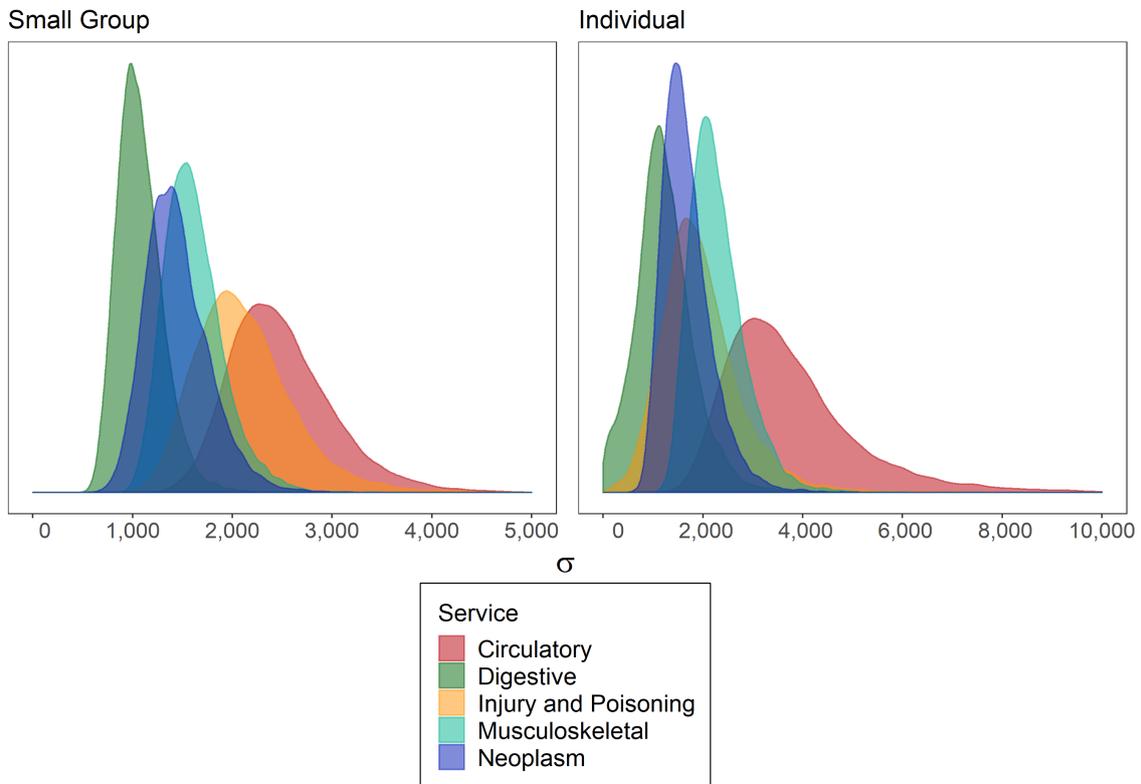


Figure B.2: MCMC Posterior densities of σ_μ (in USD) for each *Surgical and Transplant* service in the small group and individual market segments

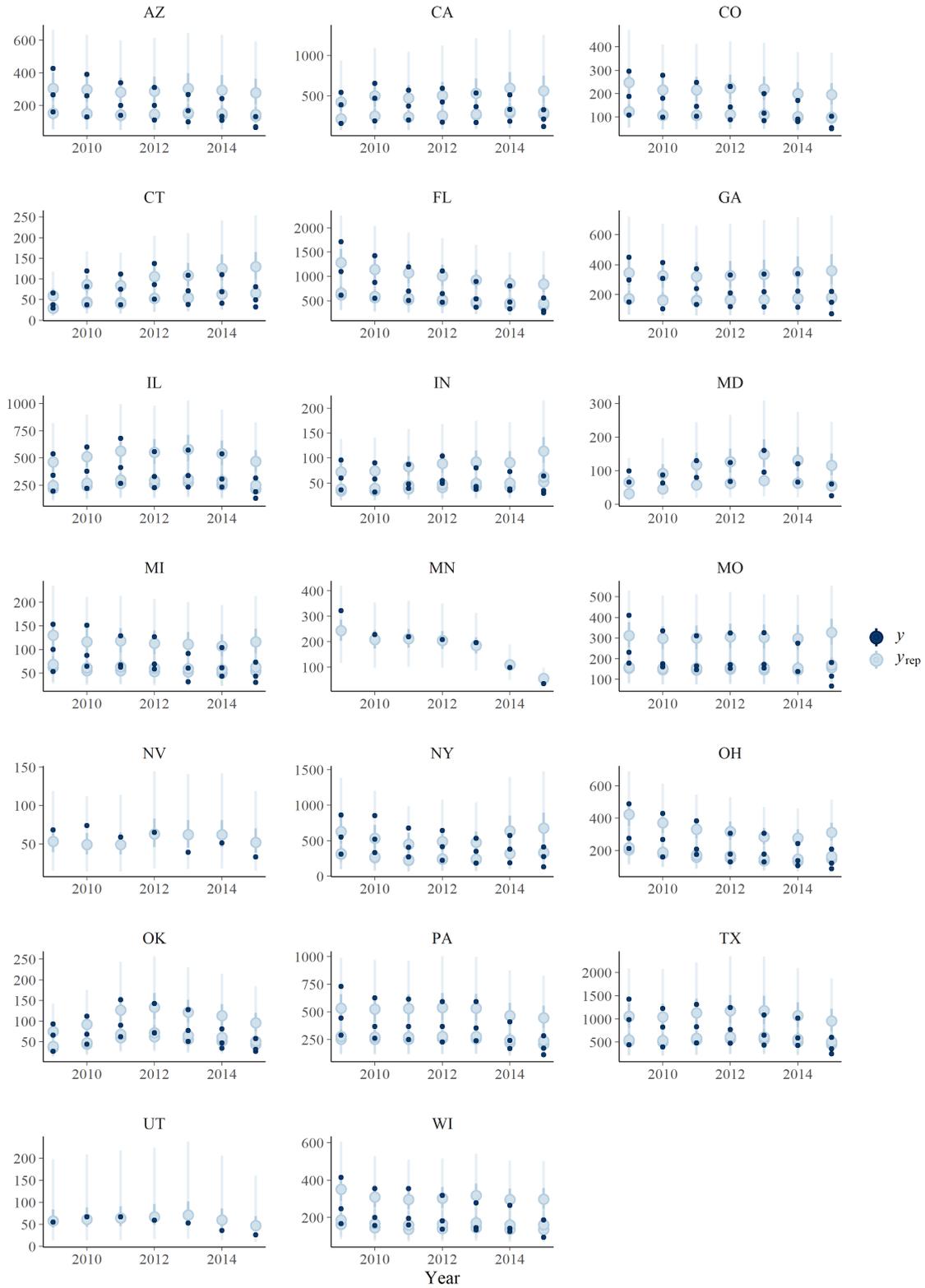


Figure B.3: *Neoplasm* claim counts 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

B.2.1 Small group

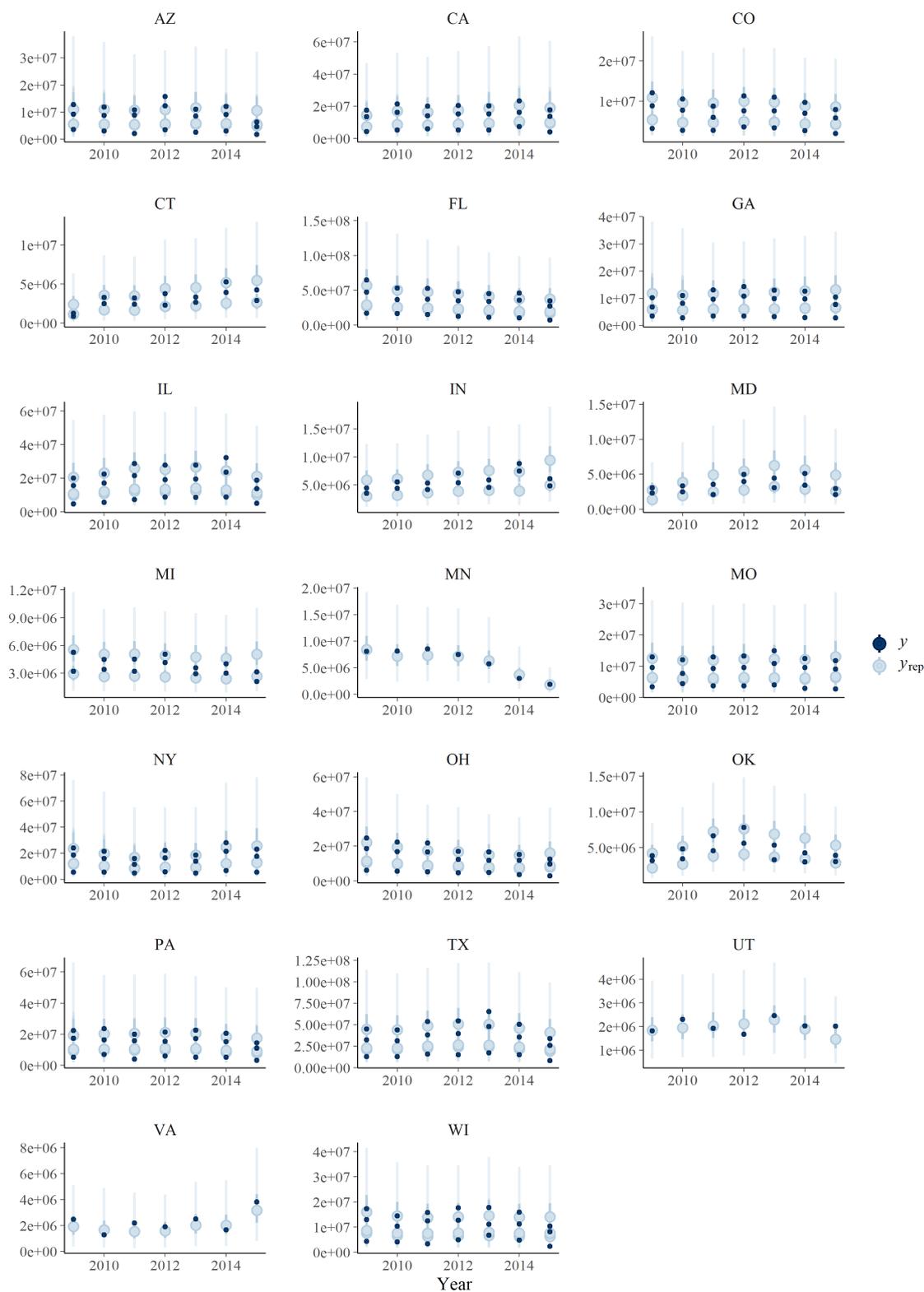


Figure B.4: *Circulatory* aggregate claims (in USD) 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

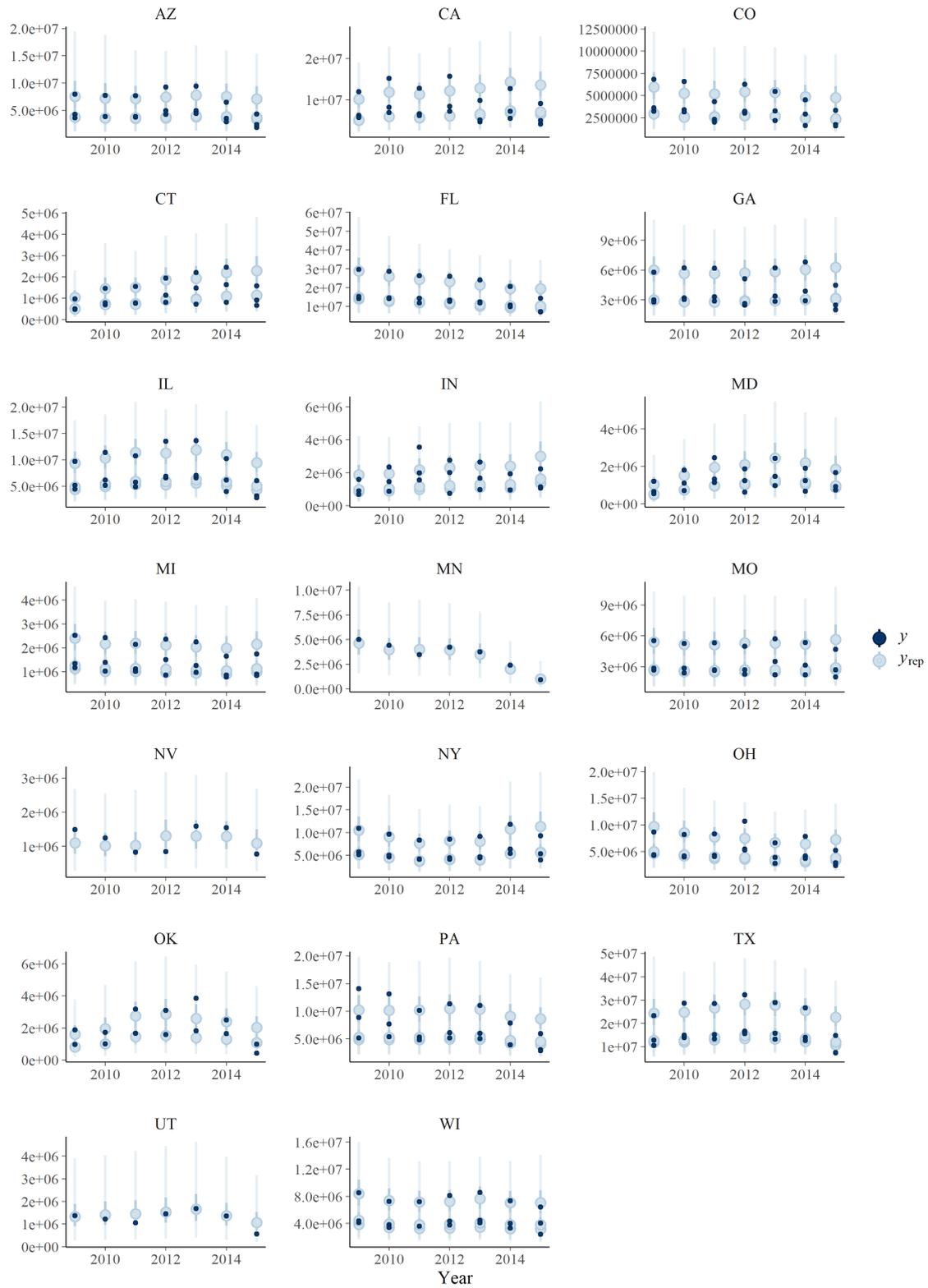


Figure B.5: *Digestive* aggregate claims (in USD) 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

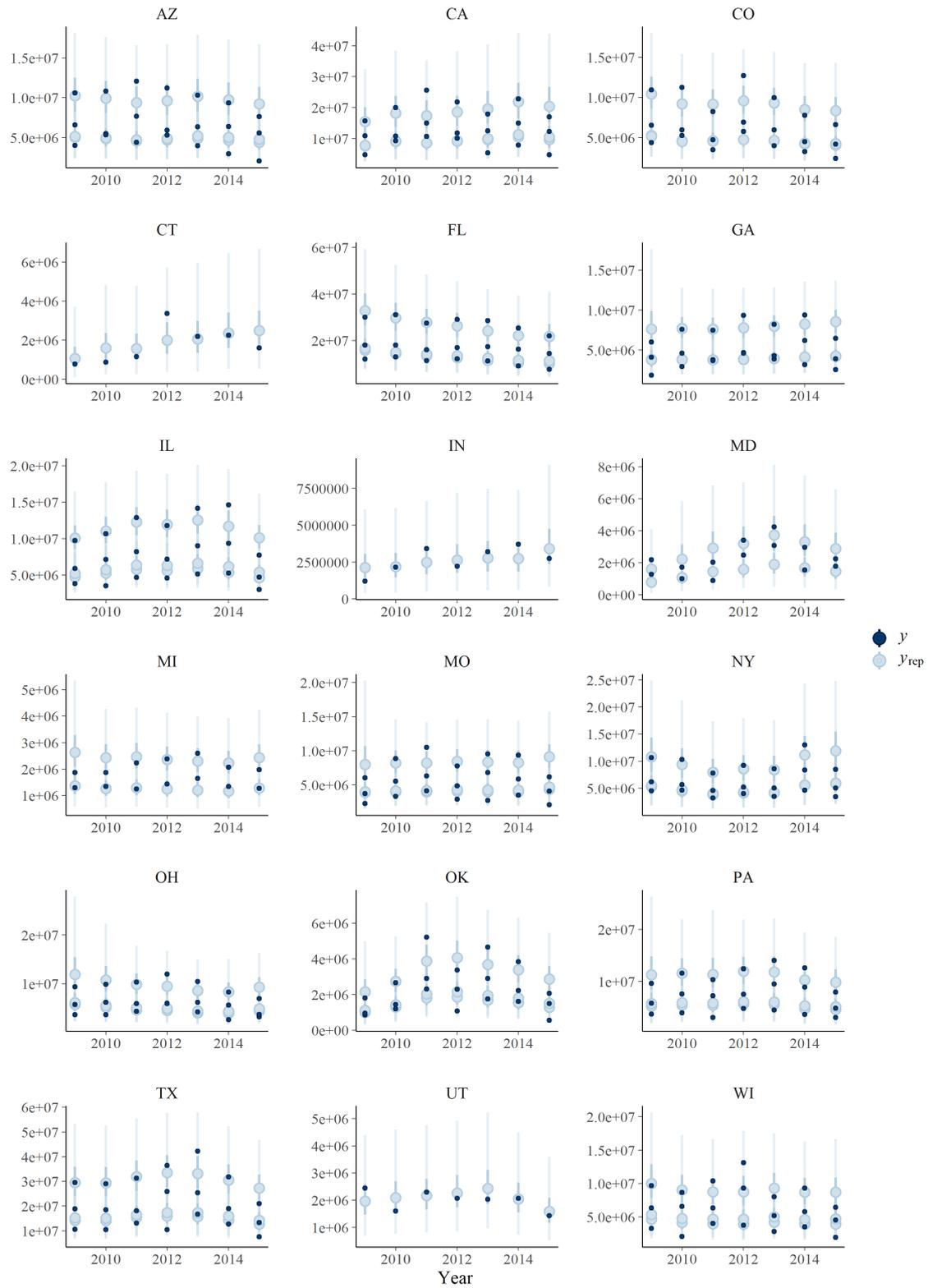


Figure B.6: *Injury and Poisoning* aggregate claims (in USD) 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male, Female* or *All*).

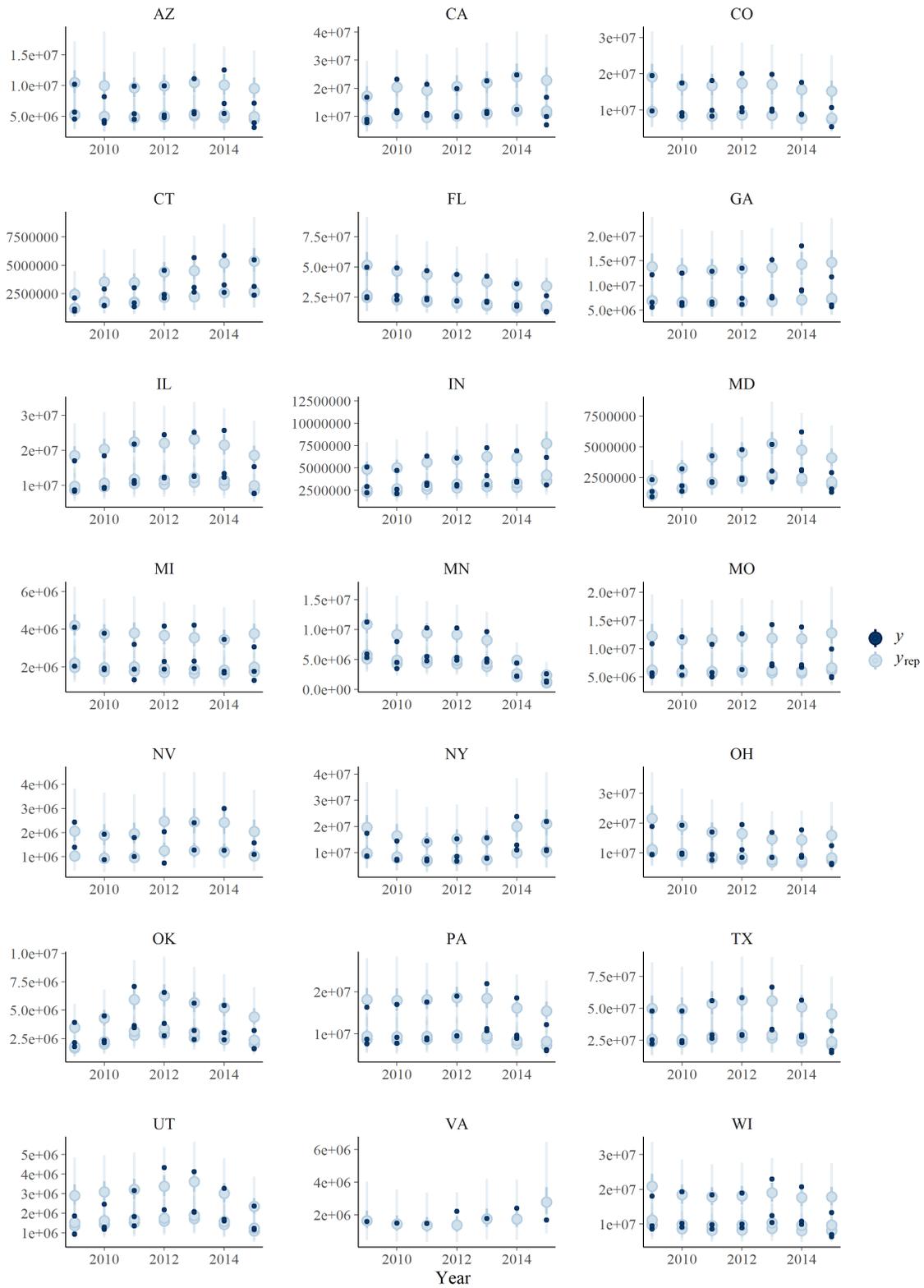


Figure B.7: *Musculoskeletal* aggregate claims (in USD) 95% credible intervals and medians (light blue) in small group markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

B.2.2 Individual

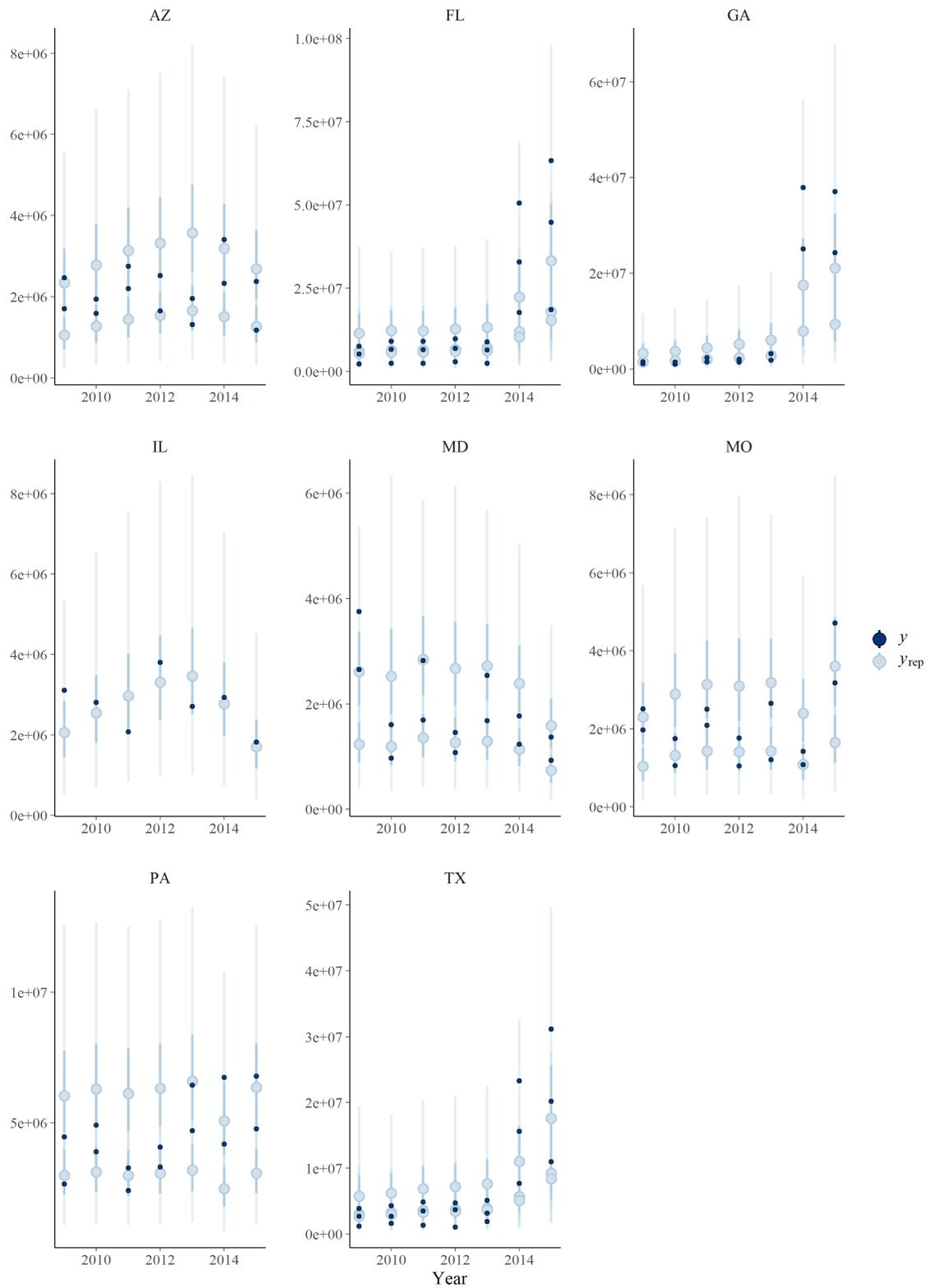


Figure B.8: *Circulatory* aggregate claims (in USD) 95% credible intervals and medians (light blue) in individual markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

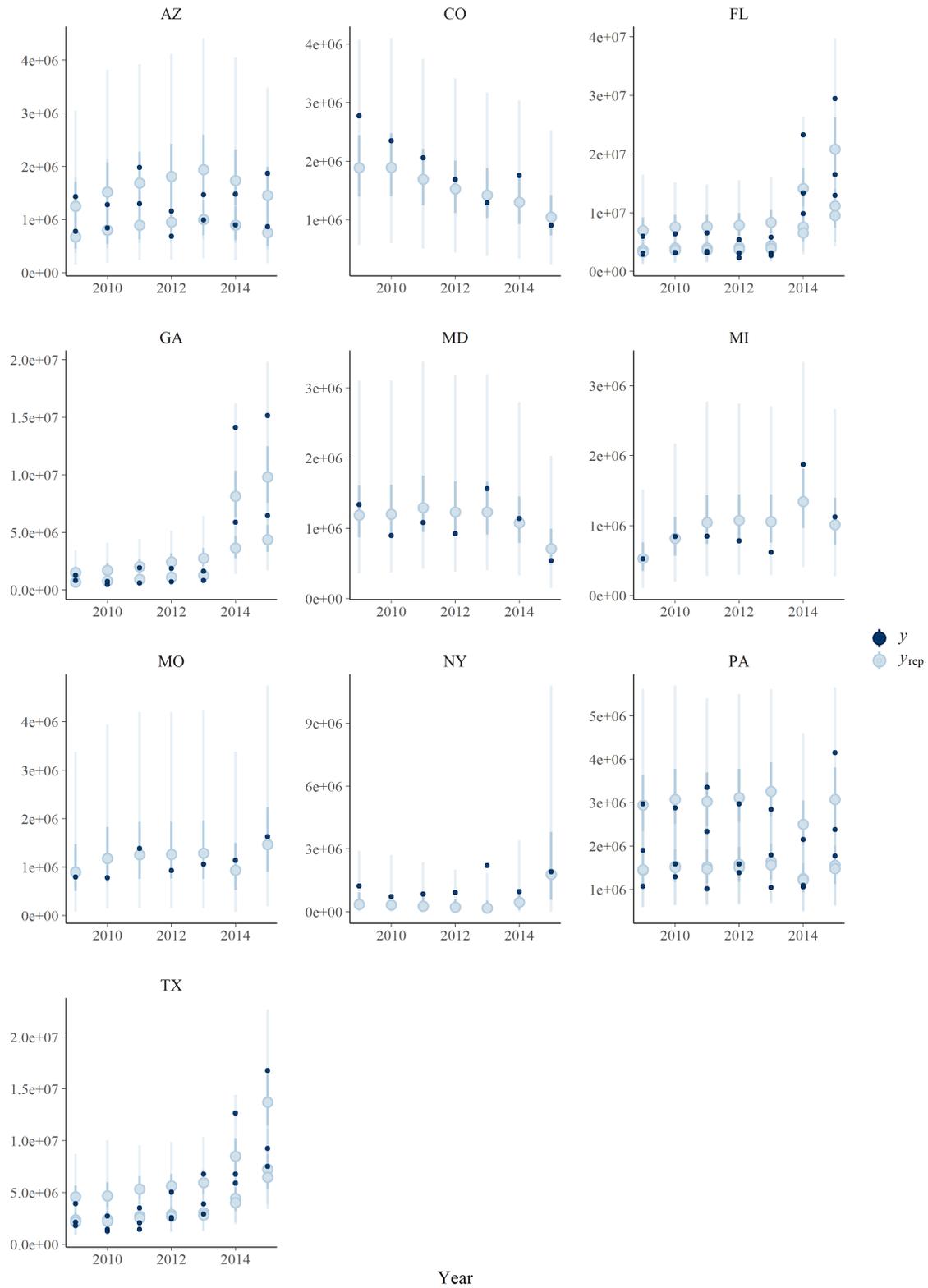


Figure B.9: *Digestive* aggregate claims (in USD) 95% credible intervals and medians (light blue) in individual markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

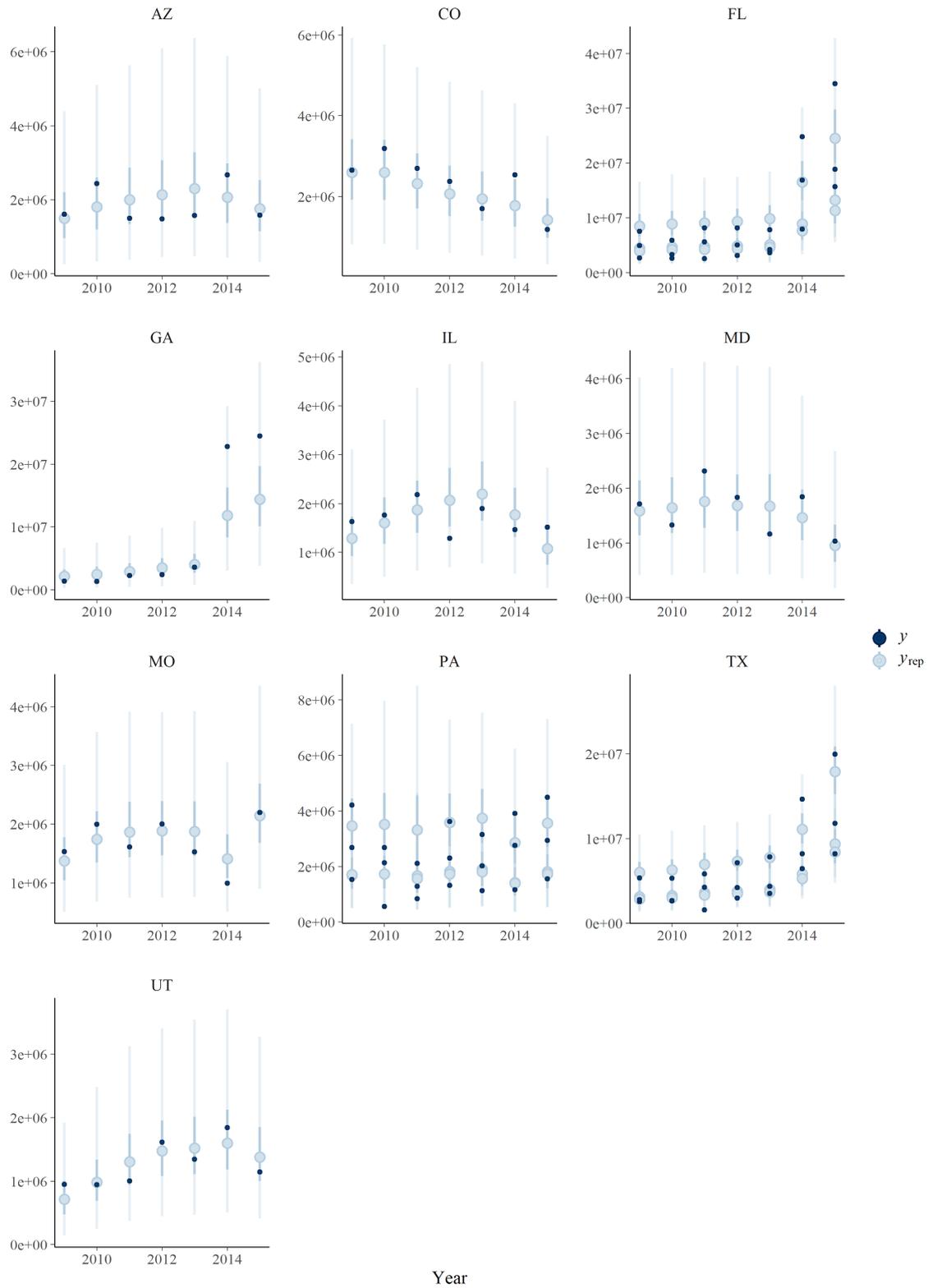


Figure B.10: *Injury and Poisoning* aggregate claims (in USD) 95% credible intervals and medians (light blue) in individual markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male, Female* or *All*).

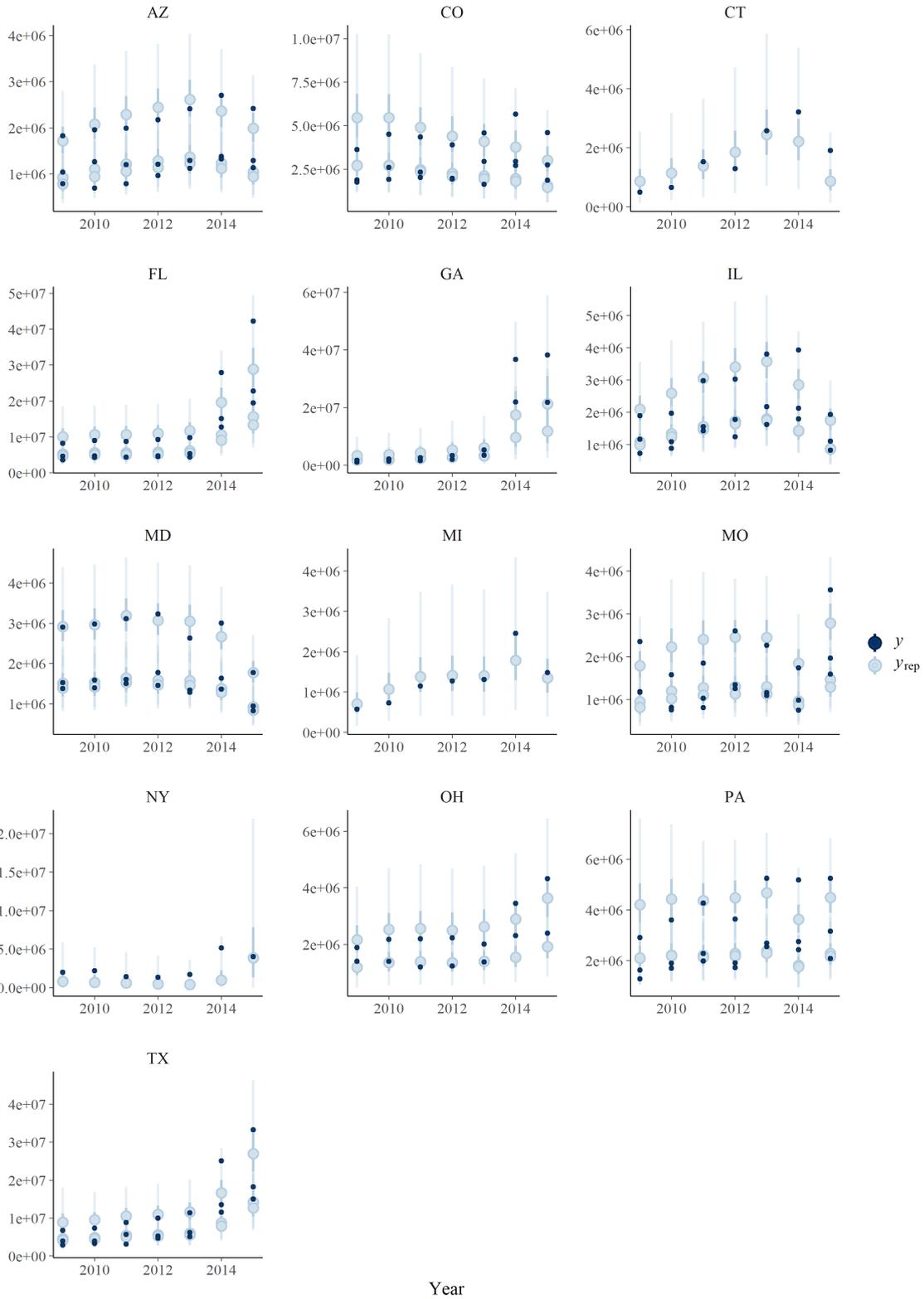


Figure B.11: *Musculoskeletal* aggregate claims (in USD) 95% credible intervals and medians (light blue) in individual markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male, Female* or *All*).

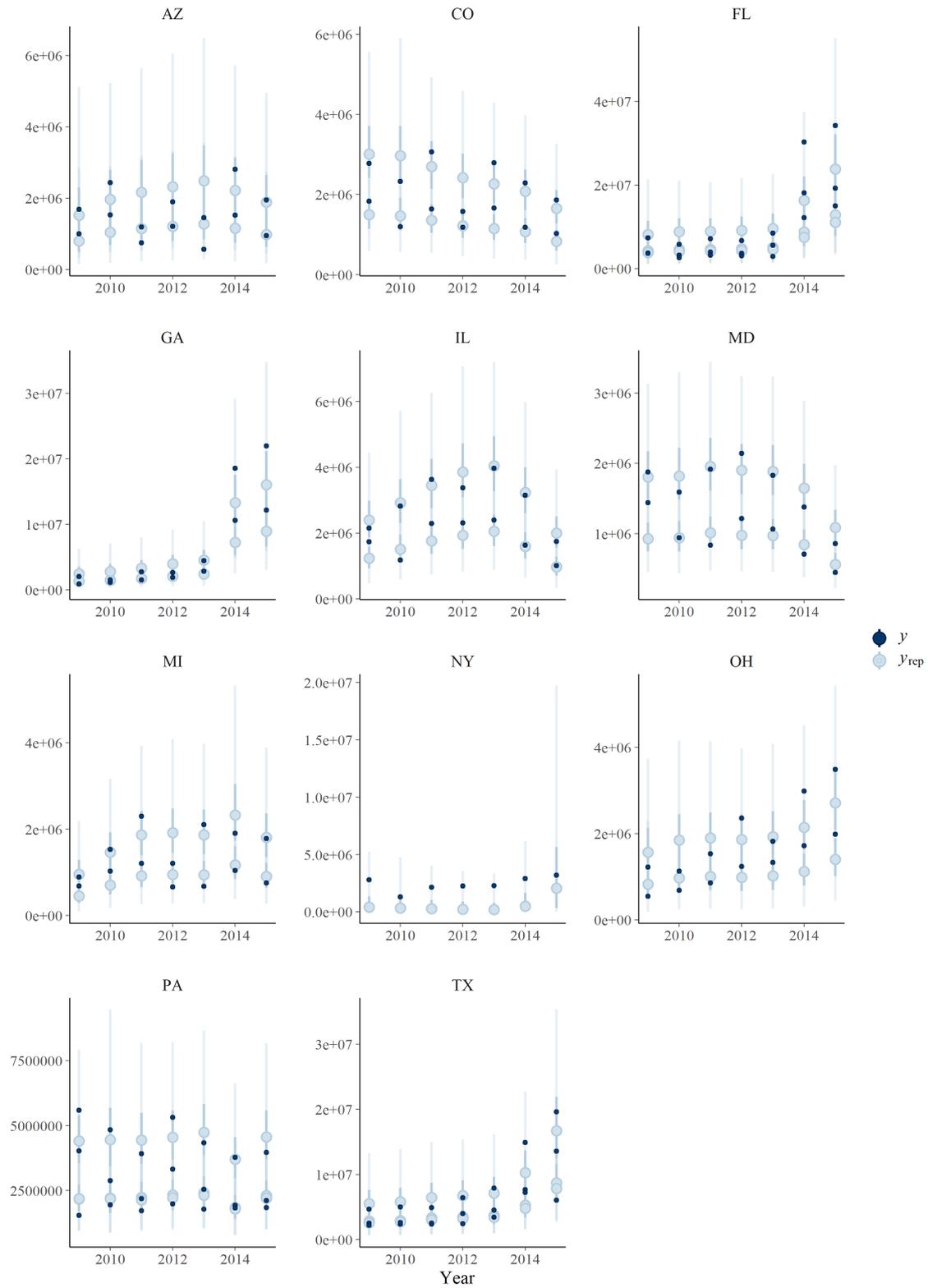


Figure B.12: *Neoplasm* aggregate claims (in USD) 95% credible intervals and medians (light blue) in individual markets for the period 2009 to 2015. Observed outcomes (dark blue) are associated to a risk pool (*Male*, *Female* or *All*).

Appendix C

Stan code

```
functions{
  real beta_prime_2_lpdf(real y, real alpha, real gamma, real theta){
    real ylpdf = -lbeta(alpha, gamma) - log(theta);
    ylpdf += (alpha - 1)*log(y/theta);
    ylpdf += -(alpha + gamma)*log1p(y/theta);
    return ylpdf;
  }
}

data {
  int<lower=0> N; //number of observations
  int<lower=0> K; //number of states
  int<lower=0> M; //number of gamma ts
  int<lower=0> J; //ts from 2014
  vector<lower=0>[N] S_tsg; //Aggregate loss of group tsg (response variable)
  int N_tsg[N]; //Total number of claims for group tsg
  vector<lower=0>[N] n_tsg ; //Auxiliar
  vector<lower=0>[N] m_tsg ; //membermonths
  vector<lower=0>[K] m_s2015 ; //membermonths 2015
  matrix[N,K] design_matrix_1; //design matrix state on level 1
  matrix[N,M] design_matrix_2; //design matrix year:state on level 1
  matrix[M,K] design_matrix_3; //design matrix state on level 2
  real<lower=0> wfreq; //observed weighted frequency
  real<lower=0> wsdfreq; //observed weighted std of frequency
  real<lower=0> wmCPS; // observed weighted CPS
  real<lower=0> wsdCPS; // observed weighted std of CPS
}

parameters {
  //parameters for S_tsg
  vector<lower=3>[M] gamma_ts;

  vector<lower=1>[K] phi_s;

  vector[K] mu_tilde;
  real<lower=0> mu;
  real<lower=0> sigma_mu;
}
```

```

//parameters N_tsg
vector<lower=0>[K] delta_s;
vector<lower=0>[K] zeta_s;
real<lower=0> sigma_zeta;
real<lower=0> zeta;

}

transformed parameters {

vector<lower=0>[K] mu_s= rep_vector(mu,K) + rep_vector(sigma_mu,K) .* mu_tilde;

vector<lower=0>[M] mu_index_2= design_matrix_3* mu_s;
vector<lower=1>[M] phi_index_2= design_matrix_3* phi_s;

vector<lower=0>[M] alpha_ts = (gamma_ts -1) ./ (phi_index_2 .* gamma_ts - 2 *
    ↪ phi_index_2 - 1);
vector<lower=0>[M] theta_ts = mu_index_2 .* (phi_index_2 .* gamma_ts - 2 *
    ↪ phi_index_2 - 1);

vector<lower=0>[N] theta_index=design_matrix_2* theta_ts;
vector<lower=3>[N] gamma_index=design_matrix_2* gamma_ts;
vector<lower=0>[N] alpha_index=design_matrix_2* alpha_ts;

}

model {
//prior specification

target += normal_lpdf(phi_s|1,25);

target += normal_lpdf(mu_tilde|0,3);
target += normal_lpdf(mu|wmCPS,wsdCPS/2);
target += cauchy_lpdf(sigma_mu|0,wsdCPS);

target += normal_lpdf(gamma_ts|3,100);

target += cauchy_lpdf(delta_s|0,1);
target += cauchy_lpdf(sigma_zeta|0,wsdfreq);
target += normal_lpdf(zeta|wfreq,wsdfreq/2);
target += normal_lpdf(zeta_s|zeta,sigma_zeta);

//target density

for(i in 1:N)
target += beta_prime_2_lpdf(S_tsg[i] | alpha_index[i] * n_tsg[i],gamma_index[i],
    ↪ theta_index[i]) ;

```

```

target += neg_binomial_lpmf(N_tsg | design_matrix_1* delta_s ,design_matrix_1 *
    ↪ delta_s ./ (m_tsg .* (design_matrix_1 * zeta_s))) ;
}
generated quantities{
  real srep[N];
  int nrep[N] = neg_binomial_rng(design_matrix_1* delta_s,design_matrix_1 *
    ↪ delta_s ./ (m_tsg .* (design_matrix_1 * zeta_s)));
  real srep2015[K];
  int nrep2015[K] = neg_binomial_rng(delta_s, delta_s ./ (m_s2015 .* zeta_s) );
  for (i in 1:N){
    if (nrep[i]==0)
      srep[i]=0;
    else
      srep[i] = theta_index[i] * gamma_rng( alpha_index[i] * nrep[i], 1) / gamma_rng(
        ↪ gamma_index[i], 1);
  }
  for (i in 1:K){
    if (nrep2015[i]==0)
      srep2015[i]=0;
    else
      srep2015[i] = theta_index[J+i] * gamma_rng( alpha_index[J+i] * nrep2015[i], 1) /
        gamma_rng( gamma_ts[J+i], 1);
  }
}

```

Bibliography

- Amin, Z. and M. Salem (2015). Bayesian modelling of health insurance losses. *Journal of Applied Statistics* 42(2), 231–251.
- Bailey, A. L. (1950). *Credibility Procedures: Laplace's generalization of Bayes' Rule and the combination of collateral knowledge with observed data*. New York State Insurance Department.
- Berger, J. et al. (2006). The case for objective Bayesian analysis. *Bayesian analysis* 1(3), 385–402.
- Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.
- Bühlmann, H. and E. Straub (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries* 70(1), 111–133.
- Centers for Medicare and Medicaid Services (2019). Potential Updates to HHS-HCCs for the HHS-operated Risk Adjustment Program. <https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/Potential-Updates-to-HHS-HCCs-HHS-operated-Risk-Adjustment-Program.pdf>. Last accessed 10 September 2020.
- Centers for Medicare and Medicaid Services (2020). FAQs about Families First Coronavirus Response Act and Coronavirus Aid, Relief, and Economic Security Act implementation Part 42. <https://www.cms.gov/files/document/FFCRA-Part-42-FAQs.pdf>. Last accessed 25 July 2020.
- Chen, S.-L. (2016). Economic costs of hemophilia and the impact of prophylactic treatment on patient management. *Am J Manag Care* 22(5 Suppl), s126–33.
- Cox, C., D. McDermott, and R. Fehr (2019). Individual insurance market performance in 2019. *Kaiser Family Foundation*.
- De Vylder, F. (1984). Practical models in credibility theory, including parameter estimation. In *Premium Calculation in Insurance*, pp. 133–150. Springer.
- Dickson, D. C. M. (2016). *Insurance Risk and Ruin* (2 ed.). International Series on Actuarial Science. Cambridge University Press.
- Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth (1987). Hybrid monte carlo. *Physics letters B* 195(2), 216–222.
- Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika* 16(1), 27–31.
- Efron, B. (1986). Why isn't everyone a bayesian? *The American Statistician* 40(1), 1–5.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (2013). *Modelling extremal events: for insurance and finance*, Volume 33. Springer Science & Business Media.

- Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics* 31(7), 799–815.
- Gao, G., Gao, and Zhang (2018). *Bayesian Claims Reserving Methods in Non-life Insurance with Stan*. Springer.
- Gelfand, A. E. and A. F. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* 85(410), 398–409.
- Gelman (2014). Discussion with Sander Greenland on posterior predictive checks. <https://statmodeling.stat.columbia.edu/2014/08/11/discussion-sander-greenland-posterior-predictive-checks/>. Last accessed 6 June 2020.
- Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics* 48(3), 432–435.
- Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis* 1(3), 515–534.
- Gelman, A. et al. (2007). Struggles with survey weighting and regression modeling. *Statistical Science* 22(2), 153–164.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. CRC press.
- Gelman, A., D. Lee, and J. Guo (2015). Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics* 40(5), 530–543.
- Gelman, A., D. B. Rubin, et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* 7(4), 457–472.
- Gelman, A. and C. R. Shalizi (2013). Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology* 66(1), 8–38.
- Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (6), 721–741.
- Gesmann, M. and J. Morris (2020). Hierarchical compartmental reserving models. In *Casualty Actuarial Society*, pp. 4.
- Goldstein, M. and D. Wooff (2007). *Bayes linear statistics: Theory and methods*, Volume 716. John Wiley & Sons.
- Grün, B., I. Kosmidis, and A. Zeileis (2011). Extended beta regression in R: shaken, stirred, mixed, and partitioned. Technical report, Working Papers in Economics and Statistics.
- Hachemeister, C. (1975). Credibility for regression models with application to trend (reprint). *Credibility: Theory and Applications*. Edited by P. Kahn. New York: Academic Press, Inc, 307–48.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), 97–109.

- Hoffman, M. D. and A. Gelman (2014). The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1), 1593–1623.
- Jewell, W. S. (1974). Credible means are exact Bayesian for exponential families. *ASTIN Bulletin: The Journal of the IAA* 8(1), 77–90.
- Jewell, W. S. (1975). The use of collateral data in credibility theory: a hierarchical model.
- Kaas, R., M. Goovaerts, J. Dhaene, and M. Denuit (2008). *Modern actuarial risk theory: using R*, Volume 128. Springer Science & Business Media.
- Kautter, J., G. C. Pope, and P. Keenan (2014). Affordable Care Act risk adjustment: overview, context, and challenges. *Medicare & Medicaid Research Review* 4(3).
- KFF (2019). Key Facts about the Uninsured Population. <https://www.kff.org/uninsured/issue-brief/key-facts-about-the-uninsured-population/>. Last accessed 10 July 2020.
- Klugman, S. (1991). *Bayesian Statistics in Actuarial Science: with Emphasis on Credibility*. Huebner International Series on Risk, Insurance and Economic Security. Springer Netherlands.
- Klugman, S. A., H. H. Panjer, and G. E. Willmot (2012). *Loss models: from data to decisions*, Volume 715. John Wiley & Sons.
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kupper, J. (1962). Wahrscheinlichkeitstheoretische modelle in der schadenversicherung. *Blätter der DGVFM* 5(4), 451–503.
- Lindley, D. V. (1975). The future of statistics: A bayesian 21st century. *Advances in Applied Probability* 7, 106–115.
- Lindley, D. V. and A. F. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)* 34(1), 1–18.
- Livingston, S. (2017). Northwell Health shutting down CareConnect insurance unit. <https://www.modernhealthcare.com/article/20170824/NEWS/170829935/northwell-health-shutting-down-careconnect-insurance-unit>. Last accessed 10 July 2020.
- Mayerson, A. L. (1964). A Bayesian view of credibility. In *Proceedings of the Casualty Actuarial Society*, Volume 51, pp. 7–23.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics* 21(6), 1087–1092.
- Meyers, G. and N. Schenker (1983). Parameter uncertainty in the collective risk model. *PCAS LXX*, 111.

- Migon, H. S. and F. A. Moura (2005). Hierarchical bayesian collective risk model: an application to health insurance. *Insurance: Mathematics and Economics* 36(2), 119–135.
- Migon, H. S. and E. M. Penna (2006). Bayesian analysis of a health insurance model.
- Natsis, Owen, H. H. (2019). Commercial Health Care Cost and Utilization Trends From 2009–2015.
- Neuhaus, W. (1984). Inference about parameters in empirical linear bayes estimation problems. *Scandinavian Actuarial Journal* 1984(3), 131–142.
- Neuhaus, W. (1995). Community rating and equalisation 1. *ASTIN Bulletin: The Journal of the IAA* 25(2), 95–118.
- Ohlsson, E. and B. Johansson (2010). *Non-life insurance pricing with generalized linear models*, Volume 2. Springer.
- Panjer, H. H. (1980). The aggregate claims distribution and stop-loss reinsurance. *Proceedings of the Casualty Actuarial Society*, XXXII.
- Panjer, H. H. and G. E. Willmot (1983). Compound poisson models in actuarial risk theory. *Journal of Econometrics* 23(1), 63–76.
- Pope, G. C., H. Bachofer, A. Pearlman, J. Kautter, E. Hunter, D. Miller, and P. Keenan (2014). Risk transfer formula for individual and small group markets under the Affordable Care Act. *Medicare & Medicaid Research Review* 4(3).
- Pupp, R. L. (1981). Community rating and cross subsidies in health insurance. *Journal of Risk and Insurance* 48, 610–627.
- Robert, C. and G. Casella (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- Rosenberg, J. (2019). FDA Approves Gene Therapy With \$2.1M Price Tag for Spinal Muscular Atrophy in Pediatric Patients. <https://www.ajmc.com/view/fda-approves-gene-therapy-with-21m-price-tag-for-spinal-muscular-atrophy-in-pediatric> Last accessed 10 July 2020.
- Sarabia, J. M., E. Gómez-Déniz, F. Prieto, and V. Jordá (2016). Risk aggregation in multivariate dependent pareto distributions. *Insurance: Mathematics and Economics* 71, 154–163.
- Smith, A. F. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)* 55(1), 3–23.
- Society of Actuaries (2019). Commercial Health Care Cost data extract. <https://www.soa.org/resources/research-reports/2019/commercial-health-care-cost-utilization-trends/>. Last accessed 10 July 2020.
- Taylor, G. C. (1974). Experience rating with credibility adjustment of the manual premium. *ASTIN Bulletin: The Journal of the IAA* 7(3), 323–336.

- Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics* 22, 1701–1728.
- Venter, G. (1983). Transformed beta and gamma distributions and aggregate losses. In *Proceedings of the Casualty Actuarial Society*, Volume 70, pp. 156–193.
- Venter, G. (1985). Structured credibility in applications—hierarchical, multidimensional, and multivariate models. *Actuarial Research Clearing House* 2, 267–308.
- Whitney, A. W. (1918). The Theory of Experience Rating. In *Proceedings of the Casualty Actuarial Society*, Volume 4, pp. 274–292.
- Zorn, J. (2016). Loss of HealthyCT is Unhealthy for Connecticut. <https://universalhealthct.org/loss-of-healthyct-is-unhealthy-for-connecticut/>. Last accessed 10 July 2020.