

# JOINT MODELING OF CLAIM FREQUENCIES AND BEHAVIORAL SIGNALS IN MOTOR INSURANCE

Alexandre Corradin, Michel Denuit, Marcin  
Detyniecki, Vincent Grari, Matteo  
Sammarmo, Julien Trufin

REPRINT | 2022 / 04

## **ISBA**

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : [lidam-library@uclouvain.be](mailto:lidam-library@uclouvain.be)

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

# JOINT MODELING OF CLAIM FREQUENCIES AND BEHAVIORAL SIGNALS IN MOTOR INSURANCE

Alexandre Corradin<sup>1</sup>, Michel Denuit<sup>2</sup>, Marcin Detyniecki<sup>1</sup>,  
Vincent Grari<sup>1</sup>, Matteo Sammarco<sup>1</sup>, Julien Trufin<sup>3</sup>

<sup>1</sup>AXA/GO/REV, Paris, France

<sup>2</sup>Institute of Statistics, Biostatistics and Actuarial Science (ISBA/LIDAM),  
UCLouvain, Louvain-la-Neuve, Belgium

<sup>3</sup>Department of Mathematics,  
Université Libre de Bruxelles (ULB), Bruxelles, Belgium

December 12, 2020

## **Abstract**

Telematics devices installed in insured vehicles provide actuaries with new risk factors, such as the time of the day, average speeds and other driving habits. This paper extends the multivariate mixed model describing the joint dynamics of telematics data and claim frequencies proposed by Denuit et al. (2019a) by replacing the estimation procedure with the Expected Conditional Maximization (ECM) algorithm. A case study performed on a database related to Pay-How-You-Drive, or PHYD motor insurance illustrates the relevance of the proposed approach for practice.

**Keywords:** Risk classification, premium calculation, driving behavior, mixed models, multivariate credibility, panel data.

# 1 Introduction

Insurers have now started to collect telematics data from usage-based motor insurance comprising high-frequency GPS location, road conditions, driving distances, duration of trips, speed, force of acceleration and deceleration or changes of direction, for instance. This provides actuaries with valuable information that can be used for ratemaking, beyond classical risk factors used for decades.

Traditionally, motor insurance pricing is generally performed in two steps:

**a priori, risk classification:** first, a priori features are integrated in the pure premium with the help of supervised learning models including policyholder’s characteristics as well as information about his or her vehicle and about the type of coverage selected, among others.

**a posteriori, experience rating:** then, a posteriori information is used to refine the a priori risk evaluation. This is done with the help of a credibility, or mixed model inducing serial dependence among past and future claims by means of random effects accounting for unexplained heterogeneity. Credibility models are sometimes simplified into bonus-malus scales for commercial purposes.

A priori means here that features are available before the start of the coverage period. This is the case for policyholder’s age, gender or place of residence, power or use of the vehicle, or policy conditions, for instance. We refer the reader to Denuit et al. (2007) for an extensive presentation of motor insurance pricing techniques.

Telematics data reflecting actual behavior behind the wheel have been studied both in actuarial and transportation science. Instead of proxies, the actual driving behavior can be captured by measuring speeding or non-fluent driving (i.e. frequent acceleration and deceleration), for instance, thanks to telematics. Here, we focus on the actuarial point of view. Because of the massive amount of information collected by telematics devices installed in the insured vehicle, many authors purposed to extract meaningful, low-dimensional signals summarizing driving experience, to be included in supervised learning models targeting claim characteristics. For instance, Boucher et al. (2013) and Ayuso et al. (2016) study the risk exposure to total driving distances and its interaction with gender and policy duration. Wuthrich (2017) suggested unsupervised classification of driving styles into a few categories that can then supplement traditional risk factors in supervised learning. Ayuso et al. (2019) and Guillen et al. (2019) study the risk factors associated with the percentages of kilometers driven above speed limits, the distances in urban area, and night driving. Verbelen et al. (2018) partitioned the total driving distances by road type and time slots and then used these compositional covariates and classical risk factors in claims frequency models. Weidner et al. (2016, 2017) investigated driving styles by studying acceleration patterns using the discrete Fourier transform. The v-a heatmaps developed by Gao et al. (2019a) after Wuthrich (2017) describe the speed-acceleration pattern as a two-dimensional functional illustrated by a pixel matrix. This reflects how drivers accelerate or brake at different speeds. Gao et al. (2019b) implemented k-medoids clustering and principal components analysis to decompose the v-a heatmap functional in different speed buckets. Guillen et al. (2020) suggested to use additional information on acceleration, braking, and cornering provided by telematics

devices to identify near-miss events, a concept taken from aviation that denotes a situation that might have resulted in an accident. Such near-miss events should be excellent predictors for claim occurrence since the occurrence of near misses is likely to be related to a higher risk of being involved in future accidents.

All these approaches have in common that they extract some new features from telematics car driving data collected by internal sensors installed in cars or by drivers' mobile phones. These features are then included in supervised learning models, supplementing the classical features used for decades in the industry. In contrast, Denuit et al. (2019a) explicitly recognize that telematics data develop over time, in parallel to claim variables, and should thus be treated accordingly. Precisely, posterior telematics signals are incorporated into a multivariate credibility model, together with claim frequencies.

Because telematics data evolve over time in parallel to claim experience being recorded while policyholders are driving, signals extracted from telematics data should be treated as a posteriori information, not as a priori risk features. In order to capture the multivariate dynamics across insurance periods, these signals must be modeled jointly with claim experience. This led Denuit et al. (2019a) to design a multivariate credibility model (or mixed model, in statistics) applying to a random vector joining telematics data and claim experience.

Compared to Denuit et al. (2019a), this paper innovates by extending the calibration of the multivariate mixed model describing the joint dynamics of telematics data and claim frequencies to signals with various formats. The fitting procedure proposed in Denuit et al. (2019a) only allows the actuary to include signals measured in the same units as the response considered in the analysis (the claim frequency, so that only signals consisting in event counts can be analyzed). In practice, signals may however have various formats. Some of them may be proportions, other ones may correspond to continuous, possibly zero-augmented measurements, for instance. Many signals are continuous since embarked devices generally produce real measures. Recording a number of events, or rounding a continuous signal in multiples of a natural unit yields integer-valued signals. Including recorded numbers of near-miss events of different types also yields integer-valued signals. It is thus important to allow for various signal formats in the multivariate credibility model.

Thus, our contribution is not about the determination of the optimal signal predicting claim occurrence but rather on the way such signals can be linked to claim experience. The inclusion of signals with different formats requires the development of an original estimation procedure because available statistical packages generally only allow for one type of response: integer, proportions, continuous measurement or zero-augmented ones. Here, we design a powerful Expected Conditional Maximization (ECM) estimation procedure to fit the model. ECM has been proposed by Meng and Rubin (1993) to escape difficulties with optimization in the M-step of EM algorithms. To this end, the M-step is split into several sub-steps so that the problem is reduced to several computationally easier, lower-dimensional optimizations. Notice that we are not the first authors to implement ECM procedures in insurance studies. We refer the reader to Fung et al. (2019) for an application to Erlang count logit-weighted reduced mixture of experts model as well as to Fung et al. (2020) who propose a transformed Gamma logit-weighted mixture of experts model for severity regression, with an application to loss reserving.

The approach proposed in this paper is illustrated on a real data set. The a posteriori

corrections to a priori premiums are calculated, depending on past experience in terms of both claim frequencies and driving behavior as reflected in the signals recorded at individual level. This demonstrates the capabilities of the approach developed in this paper to tackle with practical applications.

The remainder of this paper is organized as follows. Section 2 describes multivariate credibility models for random vectors joining signals and claim counts. The ECM algorithm is described in Section 3. Before the conclusion, Section 4 illustrates the proposed approach on a real data set, and the results are compared with those obtained according to the classical actuarial approach. Technical elements are gathered in appendix. Throughout this text, index  $i$  refers to the policyholder under consideration, index  $t$  refers to time or trip driven by policyholder  $i$ , index  $j$  refers to a priori features, index  $k$  refers to signals, and  $l$  refers to the iterations of the ECM algorithm.

## 2 Multivariate credibility model

### 2.1 Mixed Poisson model for annual claim frequencies

This part of the modeling follows the initial proposal by Denuit et al. (2019a). Consider an insurance portfolio comprising  $n$  policies observed during several periods. Let  $N_{it}$  be the number of claims reported by policyholder  $i$ ,  $i = 1, 2, \dots, n$ , during period  $t$ ,  $t = 1, 2, \dots, T_i$ . Compared to classical actuarial studies dealing with annual periods, insurers using telematics data generally work with shorter time periods, like a quarter or a month. Also,  $t$  could refer to a single trip driven by the policyholder, or even a kilometer. Whatever the level of granularity, the modeling proceeds as described next.

At the beginning of each insurance period, the actuary has at his disposal some information about each policyholder summarized into  $p$  features  $x_{itj}$ ,  $j = 1, \dots, p$ , that may evolve with  $t$ . The a priori information  $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^\top$  is recorded in the database under consideration. Resorting to standard regression (or supervised learning) machinery, this information is integrated into the prediction of the expected number of claims, or expected (absolute) claim frequency. Specifically, denote as  $d_{it}$  the exposure-to-risk, that can be the distance driven in kilometers or time spent behind the wheel in the context of telematics data. The supervised learning model then produces a score  $\eta_{it} = \eta(\mathbf{x}_{it})$  for policyholder  $i$  in period  $t$ , that is linked to the expected number of claim by means of  $\lambda_{it} = d_{it} \exp(\eta_{it}) = \exp(\ln d_{it} + \eta_{it})$ . Here, we use the canonical link function for the Poisson regression model but other link functions can also be considered. Adding  $\ln d_{it}$  to the score  $\eta_{it}$  (i.e. treating this quantity as an offset) means that the insurer's price list is expressed per kilometer or per hour driven, and varies according to traditional risk features included in the vector  $\mathbf{x}_{it}$ . The score  $\eta_{it}$  can be calibrated by means of any Poisson regression technique, ranging from basic generalized linear models (GLM) to sophisticated machine learning algorithms.

A random effect  $\Delta_i$  is added to the score  $\eta_{it}$  to recognize the residual heterogeneity of the portfolio. It is clear that  $\mathbf{x}_{it}$  alone does not provide the actuary with a perfect knowledge of policyholder's behavior behind the wheel so that  $\Delta_i$  accounts for the missing part of the score, depending on hidden information. We refer to Denuit et al. (2007) for more details about this classical construction. Here, the random variable  $\Delta_i$  captures the residual effect of all

unknown characteristics relating to policyholder  $i$ . The numbers of claims  $N_{i1}, N_{i2}, N_{i3}, \dots$  are then assumed to be independent given  $\Delta_i$ .

Precisely, the model is based on the following assumptions:

**A1** given  $\Delta_i = \delta$ , the random variables  $N_{it}$ ,  $t = 1, 2, \dots$ , are independent and conform to the Poisson distribution with mean

$$\lambda_{it} \exp(\delta) = \exp(\ln d_{it} + \eta_{it} + \delta),$$

which is henceforth denoted as  $N_{it} \sim \text{Poisson}(\lambda_{it} \exp(\delta))$ . Formally,

$$\begin{aligned} \mathbb{P}[N_{i1} = k_1, \dots, N_{iT_i} = k_{T_i} | \Delta_i = \delta] &= \prod_{t=1}^{T_i} \mathbb{P}[N_{it} = k_t | \Delta_i = \delta] \\ &= \prod_{t=1}^{T_i} \left( \exp(-\lambda_{it} \exp(\delta)) \frac{(\lambda_{it} \exp(\delta))^{k_t}}{k_t!} \right). \end{aligned}$$

**A2** at the portfolio level, the sequences  $(\Delta_i, N_{i1}, N_{i2}, \dots)$ ,  $i = 1, 2, \dots, n$ , are assumed to be mutually independent.

**A3** the random effects  $\Delta_i$  are independent, Normally distributed with zero mean and constant variance  $\sigma_\Delta^2$ .

When the canonical log link function is used in the Poisson regression model, as assumed in A1, assumption A3 amounts to using a Poisson-LogNormal model for claim counts. Contrarily to what is generally assumed in the actuarial literature, where the random effects  $\Delta_i$  are supposed to be such that  $\mathbb{E}[\exp(\Delta_i)] = 1$ , the statistical literature devoted to mixed models assumes that the random effects  $\Delta_i$  are centered. Under assumption A3, we then have  $\mathbb{E}[\exp(\Delta_i)] = \exp(\sigma_\Delta^2/2)$  according to the formula giving the mathematical expectation for the LogNormal distribution. Therefore, the latter factor has to be included in the calculation of the a priori expected number of claims (with a linear score  $\eta_{it}$ , the intercept of the regression model has thus to be modified accordingly). Formally, the a priori expected number of claims is equal to

$$\mathbb{E}[N_{it}] = \mathbb{E}[\mathbb{E}[N_{it} | \Delta_i]] = \lambda_{it} \mathbb{E}[\exp(\Delta_i)] = \lambda_{it} \exp(\sigma_\Delta^2/2).$$

The latent unobservable  $\Delta_i$  characterizes the correlation structure of the claim counts  $N_{it}$  for each policyholder  $i$  and thus allows for transferring information contained in past claims to predict the occurrence of future claims. This can be seen from the covariance structure: for  $s \neq t$ , we have

$$\begin{aligned} \text{Cov}[N_{is}, N_{it}] &= \text{Cov}[\mathbb{E}[N_{is} | \Delta_i], \mathbb{E}[N_{it} | \Delta_i]] \\ &= \lambda_{is} \lambda_{it} \exp(\sigma_\Delta^2/2) (\exp(\sigma_\Delta^2/2) - 1) \end{aligned}$$

because the conditional covariance is zero by virtue of assumption A1. The larger  $\sigma_\Delta^2$ , the more information contained in past claims experience about the occurrence of future claims. The telematics signals are then used to gain supplementary information about  $\Delta_i$ , complementing past experience.

*Remark 2.1.* If longer panels are available then the static random effects  $\Delta_i$  can be replaced with dynamic ones  $\Delta_{i1}, \Delta_{i2}, \dots$  which discount past observations according to their seniority. This is easily done by replacing  $\Delta_i$  with a random sequence  $\Delta_{i1}, \Delta_{i2}, \dots$  obeying a Gaussian process whose covariance structure accounts for the memory effect (such as AR1, for instance).

*Remark 2.2.* Other distributions than the mixed Poisson one could be considered for the claim counts  $N_{it}$ . For instance, zero-inflated Poisson or hurdle models as proposed in Boucher and Denuit (2008) or in Boucher et al. (2007) could provide interesting alternatives.

## 2.2 Behavioral variables, or signals

In order to predict the number of claims  $N_{it}$  filed by policyholder  $i$  during period  $t$ , the insurer has  $q$  signals, henceforth denoted as  $S_{ikt}$ ,  $k = 1, \dots, q$ , at its disposal. These signals summarize the information collected by means of telematics devices installed in the vehicle and bring some information about the policyholder's behavior behind the wheel during the same period. They are hopefully strongly correlated to the missing part  $\Delta_i$  of the score.

To refine risk evaluation, we now combine past claims experience with the available signals. Exactly as  $\Delta_i$  accounts for hidden information influencing claim frequencies  $N_{it}$ , each signal  $S_{ikt}$  comprises an unobservable effect  $\Gamma_{ik}$  that reflects the quality of driving being correlated to  $\Delta_i$  and random noise. It is important to realize here that signals are also influenced by traditional risk factors included in  $\mathbf{x}_{it}$  so that we need to account for this effect in model design.

We explicitly allow for signals with different formats. To fix the ideas, we assume that signals obey distributions within the Exponential Dispersion family that comprises the Normal, Gamma and Inverse-Gaussian distribution for continuous measurements, the Binomial and Poisson distribution for event counts (as well as the Negative Binomial distribution as a border case) and the Tweedie distribution for zero-augmented distributions. These distributions are henceforth referred to as ED.

Recall from Denuit et al. (2019b) that a response  $Y$  valued in a subset  $\mathcal{S}$  of the real line  $(-\infty, \infty)$  is said to possess a distribution belonging to the ED family if  $Y$  admits a probability mass function  $p_Y$  or a probability density function  $f_Y$  of the form

$$\left. \begin{array}{l} p_Y(y) \\ f_Y(y) \end{array} \right\} = \exp \left( \frac{y\theta - a(\theta)}{\phi/\omega} \right) c(y, \phi/\omega), \quad y \in \mathcal{S}, \quad (2.1)$$

where

- $\theta$  = real-valued location parameter, called the canonical parameter
- $\phi$  = positive scale parameter, called the dispersion parameter
- $\omega$  = known positive constant, called the weight
- $a(\cdot)$  = monotonic convex function of  $\theta$
- $c(\cdot)$  = positive normalizing function.

The parameters  $\theta$  and  $\phi$  are essentially location and scale indicators, extending the mean value and variance in the Normal case to the whole family of ED distributions. This is henceforth denoted as  $Y \sim \text{ED}(\theta, \phi/\omega)$ .

*Remark 2.3.* Notice that distributions outside the ED family could also be used to describe the behavior of the signals under consideration, exactly as other distributions than the mixed Poisson one could be envisaged for the number of claims. The distributions in the GAMLSS family, as described in Denuit et al. (2019b) could be relevant in that respect, for instance. We concentrate here on the ED family because these distributions are supported by the vast majority of machine learning algorithms available in computer packages.

Then, a multivariate mixed/credibility model describes the joint dynamics of claim frequencies  $N_{it}$  and related signals  $S_{ikt}$ ,  $k = 1, \dots, q$ , accounting for the availability of a priori features  $\mathbf{x}_{it}$ . Precisely, assumptions A1-A3 are supplemented with

**A4** given  $\Delta_i$ , claim counts  $N_{i1}, N_{i2}, \dots$  are independent and independent of  $\Gamma_{ij}, S_{ij1}, S_{ij2}, \dots$  for all  $j = 1, 2, \dots, q$ .

**A5** given  $\Gamma_{ik}$ , the signal  $S_{ik1}, S_{ik2}, \dots$  are independent and independent of  $\Delta_i, N_{i1}, N_{i2}, \dots$ , and

$$S_{ikt} \sim \text{ED}_k(\nu_{ikt} + \Gamma_{ik}, \phi_k/\omega_{ikt})$$

where  $\nu_{ikt} = \nu_k(\mathbf{x}_{it})$  is the score for the  $k$ th signal based on a priori features  $\mathbf{x}_{it}$  and  $\Gamma_{ik}$  is Normally distributed with zero mean and variance  $\sigma_{\Gamma,k}^2$ . Here,  $\Gamma_{ik}$  represents the additional information contained in the  $k$ th signal about claim frequencies, corrected for the effect of the features  $\mathbf{x}_{it}$ . Also, the random vector  $(\Delta_i, \Gamma_{i1}, \Gamma_{i2}, \dots, \Gamma_{iq})$  is multivariate Normally distributed with zero mean vector and variance-covariance matrix  $\Sigma$ .

**A6** given  $(\Delta_i, \Gamma_{i1}, \Gamma_{i2}, \dots, \Gamma_{iq})$ , all the observable random variables  $N_{i1}, S_{i11}, S_{i21}, \dots, S_{iq1}, N_{i2}, S_{i12}, S_{i22}, \dots, S_{iq2}, \dots$ , are mutually independent.

The ED error structure represents the noise comprised in the observed signal  $S_{ikt}$  which does not reveal anything about claim counts. From assumptions A4-A6, we see that only the random effect  $\Gamma_{ik}$  component involved in the signal  $S_{ikt}$  is relevant to predict claim frequencies. Assumptions A4-A6 postulate that the dependance between signal and claim counts is only apparent and results from missing information. If we had a complete knowledge of policyholder's characteristics, i.e. if we knew  $\Delta_i$ , then assumption A4 tells us that the signals would not be needed for pricing. Because of limited knowledge about policyholder's driving style, the insurer uses the information contained in the signals that reveals the missing elements in expected claim counts. This is why each signal is separated into three components: the effect  $\nu_{ikt}$  of the available features  $\mathbf{x}_{it}$ , the relevant information  $\Gamma_{ik}$  contained in the signal, that may explain expected claim counts beyond the available  $\mathbf{x}_{it}$ , and the random noise captured by the ED error structure. Following the standard approach in mixed models, we assume here that the random vector  $(\Delta_i, \Gamma_{i1}, \Gamma_{i2}, \dots, \Gamma_{iq})$  is multivariate Normally distributed, with zero mean vector and its covariance structure drives the corrections brought by signals in the evaluation of future expected number of claims.

The correlation  $\rho_{\Delta, \Gamma, k}$  between  $\Delta_i$  and  $\Gamma_{ik}$  can be exploited to improve the estimation of the expected number of claims by combining observed signal values with past claims history. This is because claim counts  $N_{it}$  and signals  $S_{ikt}$  are correlated by means of the pair  $(\Delta_i, \Gamma_{ik})$  of random effects. This is easily seen as follows:

$$\text{Cov}[N_{it}, S_{ikt}] = \text{Cov}[E[N_{it}|\Delta_i, \Gamma_{ik}], E[S_{ikt}|\Delta_i, \Gamma_{ik}]]$$

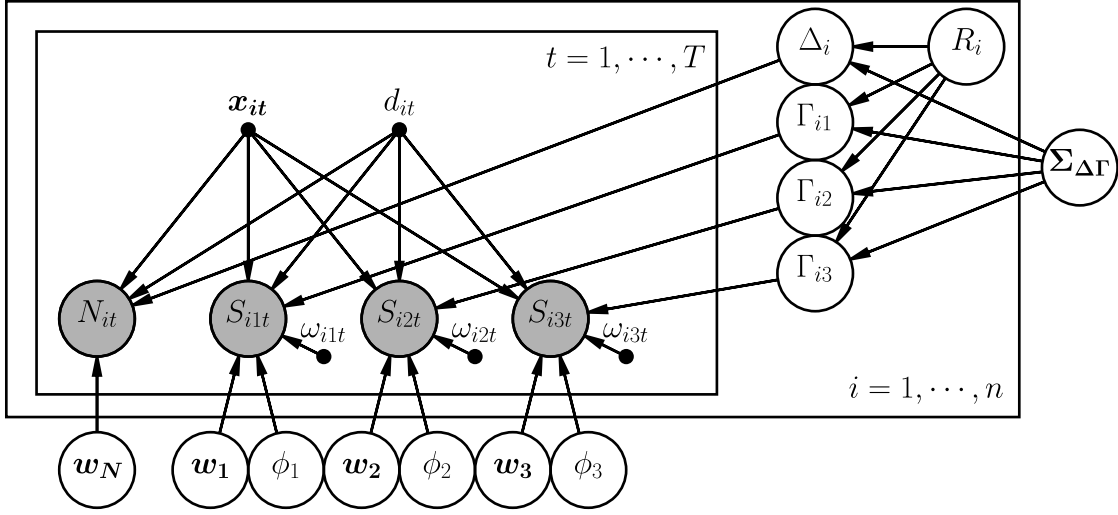


Figure 2.1: Bayesian network for the joint model of claim frequency and behavioral variables.

because the conditional covariance is zero by virtue of assumption A6. Hence, assuming that the log-link function has been used for both  $N_{it}$  and  $S_{ikt}$ , we obtain

$$\text{Cov}[N_{it}, S_{ikt}] = d_{it} \exp(\eta_{it} + \nu_{ikt}) \text{Cov}[\exp(\Delta_i), \exp(\Gamma_{ik})].$$

Now, as the pair  $(\Delta_i, \Gamma_{ik})$  is jointly Normal, with zero mean, variances  $\sigma_\Delta^2$  and  $\sigma_{\Gamma,k}^2$ , and correlation  $\rho_{\Delta, \Gamma, k}$ , we get

$$\begin{aligned} \text{Cov}[\exp(\Delta_i), \exp(\Gamma_{ik})] &= \text{E}[\exp(\Delta_i + \Gamma_{ik})] - \text{E}[\exp(\Delta_i)] \text{E}[\exp(\Gamma_{ik})] \\ &= \exp\left(\frac{\sigma_\Delta^2 + \sigma_{\Gamma,k}^2}{2}\right) \left(\exp(\rho_{\Delta, \Gamma, k} \sigma_\Delta \sigma_{\Gamma,k}) - 1\right) \end{aligned}$$

which is not zero unless  $\rho_{\Delta, \Gamma, k} = 0$ , that is, unless  $\Delta_i$  and  $\Gamma_{ik}$  are mutually independent (so that the  $k$ -th signal brings no information about the claim counts).

*Remark 2.4.* We acknowledge here that the multivariate Normal assumption may appear to be restrictive in some applications because it constrains the dependence structure (prohibiting tail dependence, for instance). Other multivariate distributions, such as Elliptical ones can be useful to model the dependency of the signals, and a copula construction can be employed to this end. The present paper describes a general modeling strategy that can be adapted to any other distributional choice.

*Remark 2.5.* The joint model of claim frequency and telematics signals can be represented by the Bayesian network displayed in Figure 2.1. This graphical display helps to understand the relationship between the different components of the proposed model and the way correlation is induced at individual level. A random variable  $R_i$  is added to the Bayesian network, it represents the level of unexplained risk of a given policyholder – in the model,  $R_i = \Delta_i$ .

### 3 ECM algorithm

Denuit et al. (2019a) considered signals obeying Mixed Poisson distributions with correlated random effects so that the `glmer` function included in the R package `lme4` can be used to fit a GLMM. To achieve convergence, some care was nevertheless needed in relation with the nonlinear optimizer. To ensure numerical stability of the optimization algorithms, some features and signals had to be rescaled, which resulted in a loss of information. Another limitation of the `glmer` function is that all signals must be mixed Poisson distributed (because signals have to follow the same law as the response, here the number of claims). For all these reasons, a more powerful estimation procedure is needed and the ECM approach proposed in Meng and Rubin (1993) offers an interesting alternative.

The expression for the likelihood of a mixed-effects model involves an integral over all the random effects. In our case, the likelihood associated to the observations  $(n_{it}, s_{i1t}, \dots, s_{iqt})$ ,  $i = 1, \dots, n$  and  $t = 1, 2, \dots, T_i$ , writes

$$\mathcal{L} = \prod_{i=1}^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \left( \exp(-d_{it} \exp(\eta_{it} + \delta)) \frac{(d_{it} \exp(\eta_{it} + \delta))^{n_{it}}}{n_{it}!} \prod_{k=1}^q f_{S_{ikt}|\Gamma_{ik}=\gamma_k}(s_{ikt}) \right) \times f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q$$

where  $f_{S_{ikt}|\Gamma_{ik}=\gamma_k}$  is the probability density function or the probability mass function of the  $k$ -th signal  $S_{ikt}$  given  $\Gamma_{ik} = \gamma_k$ . The latter has been assumed to belong to the ED family. Also,  $f_{\Sigma}$  is the joint probability density function of the random vector  $(\Delta_i, \Gamma_{i1}, \dots, \Gamma_{iq})$ , corresponding to the assumed multivariate Normal distribution with zero mean vector and variance-covariance matrix  $\Sigma$ . Clearly, a direct maximization of  $\mathcal{L}$  appears to be an extremely difficult task. This is why we follow the ECM strategy.

The EM algorithm proposed by Dempster et al. (1977) is a powerful tool to deal with mixed models because its maximum, or M-step corresponds to maximum likelihood estimations performed with complete data, after each random effect has been replaced with its conditional expectation, given observed data. This can often be performed with the help of available computer packages. Broadly speaking, the E-step can be viewed as creating a complete-data problem by imputing missing values, and the M-step can be understood as conducting a maximum likelihood-based analysis. Various methods for accelerating EM have been proposed in the literature. We refer the reader to Varadhan and Roland (2008) for a review.

The ECM algorithm proposed by Meng and Rubin (1993) replaces a complicated M-step of EM with several computationally simpler CM-steps or conditionally M-steps. Precisely, the  $l$ -th iteration of ECM consists of an E-step, which computes the expected complete-data log-likelihood function given the observed data and the current estimate of the parameter, and replaces the M-step of EM with a sequence of simpler constrained or conditional maximization (CM) steps, each of which fixes some function of the unknown parameter. Broadly speaking, ECM divides the M-step of standard EM algorithms into several sub-steps and optimize a more mathematically tractable function over a lower dimensional space in each sub-step. In the ECM approach, the parameters can be estimated using functions for fitting

GLMMs that are readily available in standard statistical software packages. Effective Gauss-Hermite quadratures are used to approximate intractable integrals. The ECM algorithm is fully described in Appendix A.

As in the EM case, the E-step consists in computing the expectation of the unknown random effects  $\Delta_i, \Gamma_{i1}, \dots, \Gamma_{iq}$ , given the observation  $N_{it}, S_{i1t}, \dots, S_{iqt}$ ,  $t = 1, \dots, T_i$ . Including these values in the offsets of the marginal models for the observations  $N_{it}, S_{i1t}, \dots, S_{iqt}$ , the scores  $\eta(\mathbf{x}_{it})$  and  $\nu_k(\mathbf{x}_{it})$  can be estimated by maximum-likelihood (CM1-step). The covariance structure of the random effects  $\Delta_i, \Gamma_{i1}, \dots, \Gamma_{iq}$  can also be estimated by maximum-likelihood on the basis of the expectations produced in the E-step (CM2-step). These three steps, the E-step followed with CM1- and CM2-steps, are iterated until convergence (see next section for the stopping criterion). The implementation of the ECM algorithm requires the numerical evaluation of conditional expectations by quadrature formulas in the E-step. Detailed explanations can be found in Appendix B.

## 4 Case study

### 4.1 Data set

To evaluate the capabilities of the multivariate credibility model presented in Section 2, we employ a data set collected by a French insurance company within the framework of a Pay-How-You-Drive motor insurance cover targeting young drivers. Data have been collected from OBD-II dongle devices. This is one of the most common way to collect telematics data among a large variety of possibilities, as reported by Ortiz et al. (2020). OBD-II dongles periodically record GPS locations as well as a time stamp and vehicle speed. From this limited, yet significant, set of measures, many other metrics can be derived, such as the number of dangerous events due to hazardous maneuvers or harsh accelerations or the kilometers driven in a given period of time as well as information as the number of kilometers driven at night or the number of kilometers driven in an urban area. Motor insurance premiums can then be computed taking into account car usage and driver's behavior behind the wheel as well as traditional risk factors.

This recorded data enables us to derive the exposure factor  $d_{it}$  that is taken to be the total driving duration.

In the following model, we consider two signals which reflect the type of driving habits and skills as follows:

$$\begin{aligned} S_{i1t} &= \text{number of dangerous driving events for policyholder } i \text{ in period } t \\ S_{i2t} &= \text{average speed for policyholder } i \text{ in period } t. \end{aligned}$$

Signal  $S_{i1t}$  is integer-valued while  $S_{i2t}$  assumes strictly positive values. Average speed values have been normalized.

The first signal mainly focuses on harsh breaking, strong accelerations. As reported by Tselentis et al. (2016), these are key factors for detecting aggressive and dangerous driving behavior. Precisely, an acceleration  $a$  is detected as a dangerous event if for at least two consecutive GPS points  $a > 3g$ . Inversely, if  $a < 3.5g$  for at least two consecutive GPS

|                 | Q1    | Q2    | Q3    |
|-----------------|-------|-------|-------|
| <b>Signal 1</b> |       |       |       |
| Min             | 0     | 0     | 0     |
| Mean            | 12.04 | 12.81 | 13.54 |
| Median          | 7     | 8     | 9     |
| IQR             | 12    | 14    | 14    |
| Max             | 178   | 172   | 291   |
| <b>Signal 2</b> |       |       |       |
| Min             | 0.08  | 0.08  | 0.07  |
| Mean            | 0.38  | 0.38  | 0.39  |
| Median          | 0.37  | 0.37  | 0.38  |
| IQR             | 0.18  | 0.19  | 0.19  |
| Max             | 1     | 1     | 1     |

Table 4.1: Sample statistics for raw telematics information by quarter ( $n = 10,446$ ). For each signal, the smallest observed value (Min), average, median, inter-quartile range (IQR) and largest observed value (Max) are displayed separately for the three quarters of observation (denoted respectively as Q1, Q2, and Q3).

readings, then a harsh braking event is raised. Acceleration events on insertion lanes are removed as well as braking events on exit lanes.

Let us now turn to the second signal included in our analysis. Average speed is an excellent proxy for road type. A high average speed will suggest that the policyholder drive in peri-urban or rural areas whereas a driver with a low average speed is very likely to drive in urban areas. We note that urban areas are often congested and crash risk is higher there than in sub-urban or rural zones, because of heavy traffic. This is exactly the information captured by the second signal.

Finally, Signals 1 and 2 are practicable and convenient to collect. In fact, apart from OBD-II dongles, the same data can be collected by the drivers’ smartphone in the context of smartphone-based motor insurance as conceived by Wahlstrom et al. (2015).

## 4.2 Descriptive statistics

The sample is made up of  $n = 10,446$  insured drivers followed over the three quarters of calendar year 2019. All policyholders have been observed for three quarters (so that  $T_i = 3$  for all  $i$ ). Thus, we have a multivariate, balanced panel structure. The maximum age for all drivers in the sample is 26. In the participating insurance company, the policies that involve collecting telematics information are only offered to young drivers.

In Table 4.1, we present descriptive statistics for telematics data comprised in the data set, separately for each quarter. It is worth stressing that Signal 2 remains remarkably stable over the observation periods, showing that driving habits in terms of speed do not change much at aggregate level. Signal 1 appears to be more volatile as driving events are more transient. It also exhibits a moderate increasing trend within the database.

Figure 4.1 displays two histograms of the raw telematics data. This helps to figure out

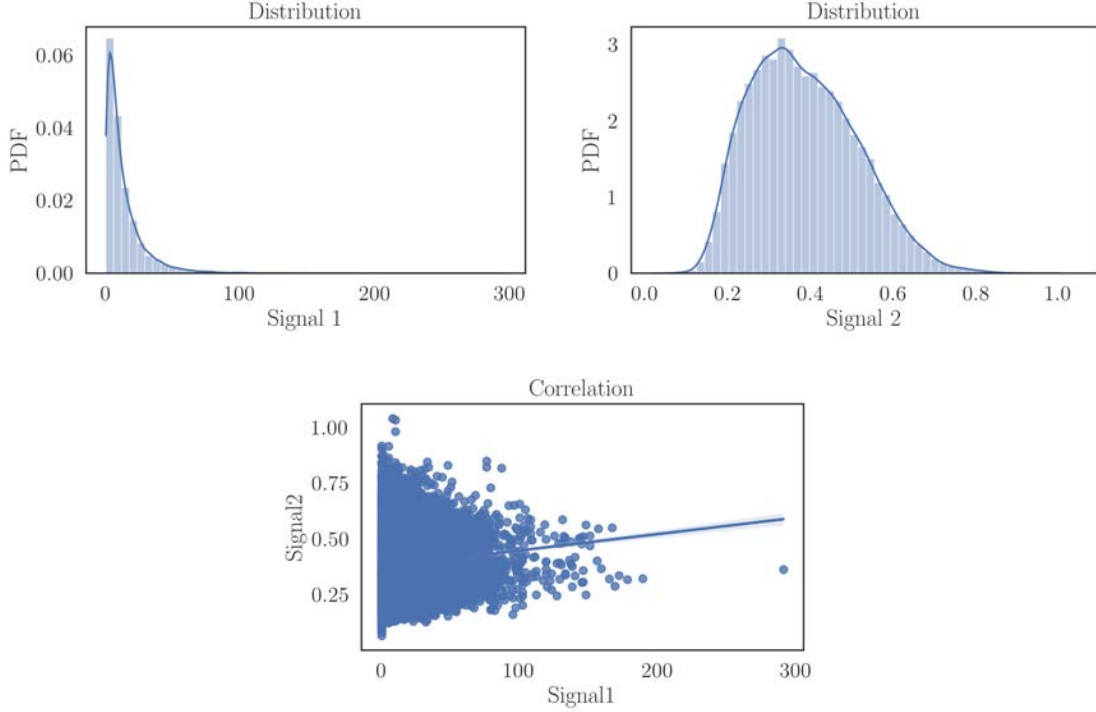


Figure 4.1: Histograms for Signals 1 and 2 (upper left and right panels, respectively) and scatterplot (bottom panel) with the linear regression line.

the sample distribution of the two signals entering the analysis. The scatterplot shows the correlation between them that appears to be moderately positive.

In order to illustrate the treatment of a priori features, we consider here two binary covariates  $x_1$  and  $x_2$  ( $p = 2$ ). The first binary feature refers to driving experience (time elapsed since the driving license has been awarded) and the second one to the power of the vehicle. They both partition policyholders in groups of roughly equal size. Table 4.2 displays the observed number of claims, together with the corresponding exposures, and the average signal values for each level of the two features. We can see there that claim frequencies as well as signal values are impacted by the features  $x_1$  and  $x_2$  so that we must enter this information in the responses to correct apparent correlation for the confounding effect of the covariates.

### 4.3 Association between signals and claim counts

We focus specifically on the two signals presented in Section 4.1 because we expect some association between claim frequencies and numbers of dangerous driving events and speed. There is an extensive literature on how all these factors are associated to claiming. Ayuso et al. (2016, 2019) showed that, among other metrics, information on speed improves the prediction of the number of claims, compared to classical models not using telematics information. Guillen et al. (2019) provide an extended overview on how accumulated distance driven shows evidence that drivers improve their skills, a phenomenon that is known as the

| Observed claim numbers<br>(with rescaled exposures) |                 |                  |
|-----------------------------------------------------|-----------------|------------------|
|                                                     | $x_2 =$         |                  |
|                                                     | 0               | 1                |
| $x_1 =$ 0                                           | 97<br>(4744.15) | 214<br>(8963.68) |
| 1                                                   | 67<br>(5051.20) | 91<br>(5765.08)  |
| Average value for Signal 1                          |                 |                  |
|                                                     | $x_2 =$         |                  |
|                                                     | 0               | 1                |
| $x_1 =$ 0                                           | 12.72           | 13.16            |
| 1                                                   | 12.08           | 12.90            |
| Average value for Signal 2                          |                 |                  |
|                                                     | $x_2 =$         |                  |
|                                                     | 0               | 1                |
| $x_1 =$ 0                                           | 0.39            | 0.36             |
| 1                                                   | 0.42            | 0.39             |

Table 4.2: Observed number of claims, together with the corresponding exposures appearing within brackets, and the average signal values for each level of the binary features  $x_1$  and  $x_2$ .

“learning effect”. All this previous knowledge is the reason why we focus specifically on variables that reflect the driving habits, such as the average speed, and for which we expect a clear association with the number of claims like the number of dangerous driving events.

Let us now investigate the strength of this association on our data set. Figure 4.2 displays the frequency of claiming policyholders within each of the 20 buckets in which Signals 1 and 2 are quantized. We clearly see there that the majority of drivers having reported claims concentrate in the higher buckets for Signal 1 and in the lower buckets for Signal 2 so that the association between signals and claim frequencies seems to be present in the data set under consideration.

We also assess the strength of non-linear dependence between the signals and the claim counts with the help of the Hirschfeld-Gebelein-Renyi (HGR) maximal correlation coefficient proposed by Grari et al. (2020). The HGR coefficient is equal to 0 if the two random variables are independent. If they are perfectly dependent the value is 1. Note that although the claim counts is integer-valued, the HGR still captures non-linear dependence since it corresponds to Chi-Square divergence in this specific case. We obtain an estimated HGR coefficient of 0.048 between claim count and Signal 1. For Signal 2 the estimation is 0.030. Signal 1 seems therefore to capture more information about the claim counts than Signal 2. Notice that the non-linear dependence between Signals 1 and 2 is 0.200.

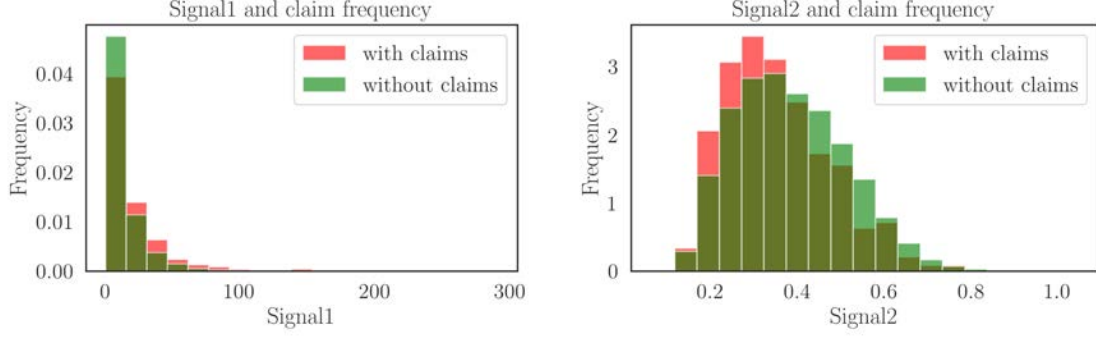


Figure 4.2: Signals 1 and 2 and claim frequency.

#### 4.4 Fitted models

We have to estimate the scores  $\eta$  for claim numbers and  $\nu_k$ ,  $k = 1, 2$ , for Signals 1 and 2, including the two features  $x_1$  and  $x_2$ , as well as dispersion parameters  $\sigma_\Delta^2$  for claim numbers,  $\sigma_{\Gamma,1}^2$  for Signal 1, and  $\phi_2$  and  $\sigma_{\Gamma,2}^2$  for Signal 2. These are the marginal parameters involved in the distribution of the number of claims and of the two signals included in the analysis. Then, the off-diagonal elements of the variance-covariance matrix  $\Sigma$  joining the random effects in claim counts and in the available signals must also be estimated, to allow information transfer from signal to claim frequencies.  $S_{i1t}$  is modelled using a Poisson regression with its natural link function while  $S_{i2t}$  is modelled using a Gamma regression using the logarithm link function.

Marginal parameters are generally well estimated by regression models fitted separately to claim frequencies and signal values. This is why the stopping criterion generally concentrates on the variance-covariance matrix  $\Sigma$  joining the claim counts to the available signals. Several distances are available to assess the proximity of two matrices. Some of them have been designed specifically for variance-covariance matrices, such as Bhattacharyya and Kullback-Leibler distance between two Gaussian distributions having the same location vector. In this paper, we use the following simple rule: we decide to stop the ECM algorithm as soon as the maximum relative difference between correlation coefficients  $\rho_{\Delta,\Gamma,k}$  at two successive steps become smaller than a pre-defined tolerance level.

Figures 4.3 and 4.4 display the estimations with respect to the number of iterations. After about 50 iterations, the estimations stabilize and the stopping criterion is fulfilled. We end up with the following estimates for the scores

$$\begin{aligned}\hat{\eta}(\mathbf{x}) &= -3.912 - 0.423x_1 + 0.155x_2 \\ \hat{\nu}_1(\mathbf{x}) &= 2.557 - 0.025x_1 - 0.134x_2 \\ \hat{\nu}_2(\mathbf{x}) &= -0.989 + 0.103x_1 - 0.073x_2\end{aligned}$$

involved in the number of claims  $N_{it}$ , Signal 1, and Signal 2, respectively. The variance-covariance matrix of the multivariate Normal distribution for the random vector  $(\Delta, \Gamma_1, \Gamma_2)$

is estimated to

$$\hat{\Sigma} = \begin{pmatrix} 0.835608624 & 0.01087129 & -0.003205314 \\ 0.010871290 & 0.54745657 & 0.020714447 \\ -0.003205314 & 0.02071445 & 0.092999400 \end{pmatrix}.$$

In accordance with the exploratory analysis, we get a negative correlation between  $\Delta$  and  $\Gamma_2$  and a positive correlation between  $\Delta$  and  $\Gamma_1$ . Despite the relatively small estimated correlations, signals induce strong a posteriori corrections, as shown next (recall from Section 2 that correlations between responses involve the exponential transform of the elements in  $\hat{\Sigma}$ ).

## 4.5 Credibility updating formulas

In the classical actuarial approach based on claim counts, only, past numbers of claims enter the credibility formulas in addition to observable features  $\mathbf{x}_{i,T_i+1}$  to explain  $N_{i,T_i+1}$ . Formally, the experience used to revise future premiums relates to past claims history, only, and is henceforth denoted as

$$\mathcal{H}_{i,T_i}^{\text{claim}} = \{N_{it}, t = 1, \dots, T_i\}.$$

This information enters the predictive distribution, i.e. the conditional distribution of  $N_{i,T_i+1}$  given  $\mathcal{H}_{i,T_i}^{\text{claim}}$ . With experience rating, the a priori expectation

$$\mathbb{E}[N_{i,T_i+1}] = \lambda_{i,T_i+1} \mathbb{E}[\exp(\Delta_i)]$$

is replaced with the a posteriori expectation

$$\mathbb{E}[N_{i,T_i+1} | \mathcal{H}_{i,T_i}^{\text{claim}}] = \lambda_{i,T_i+1} \mathbb{E}[\exp(\Delta_i) | \mathcal{H}_{i,T_i}^{\text{claim}}].$$

The pricing structure is slow to adapt in personal lines because the expected claim frequencies are generally small, whatever the driver's risk profile.

The approach proposed in this paper recognizes the a posteriori nature of telematics data. The multivariate credibility model developed in the present case study captures the association between Signals 1-2 and claim counts, allowing the actuary to refine risk evaluations based on past history. With telematics, past claims history  $\mathcal{H}_{i,T_i}^{\text{claim}}$  is enriched with behavioral data. This allows the pricing structure to become more reactive. In this case, the policy-specific history  $\mathcal{H}_{i,T_i}$  gathers all the a posteriori information

$$\mathcal{H}_{i,T_i} = \mathcal{H}_{i,T_i}^{\text{claim}} \cup \mathcal{H}_{i,T_i}^{\text{signals}} = \{(N_{it}, S_{i1t}, S_{i2t}), t = 1, \dots, T_i\}.$$

The multivariate mixed/credibility model describes the joint dynamics of  $(N_{it}, S_{i1t}, S_{i2t})$ , given a priori features  $\mathbf{x}_{it}$ . The predictive distribution now corresponds to the conditional distribution of  $N_{i,T_i+1}$  given  $\mathcal{H}_{i,T_i}$ . The a priori expectation is replaced with an a posteriori one

$$\mathbb{E}[N_{i,T_i+1} | \mathcal{H}_{i,T_i}] = \lambda_{i,T_i+1} \mathbb{E}[\exp(\Delta_i) | \mathcal{H}_{i,T_i}],$$

so that the factor  $\mathbb{E}[\exp(\Delta_i) | \mathcal{H}_{i,T_i}] / \mathbb{E}[\exp(\Delta_i) | \mathcal{H}_{i,T_i}^{\text{claim}}]$  corresponds to the improvement we get for the a posteriori correction by also using the behavioral data provided by telematics.

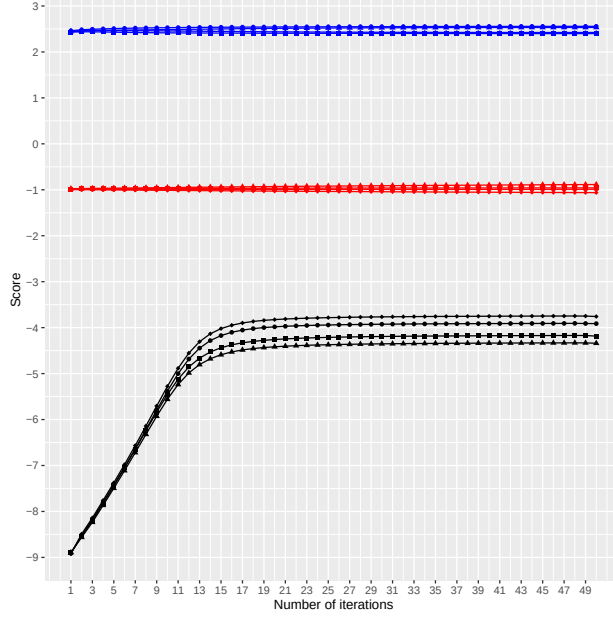


Figure 4.3: Scores estimates  $\hat{\eta}(\mathbf{x})$  (in black),  $\hat{\nu}_1(\mathbf{x})$  (in blue) and  $\hat{\nu}_2(\mathbf{x})$  (in red) along algorithm iterations for  $\mathbf{x} = (0,0)$  (circle),  $\mathbf{x} = (1,0)$  (triangle),  $\mathbf{x} = (0,1)$  (diamond) and  $\mathbf{x} = (1,1)$  (square).

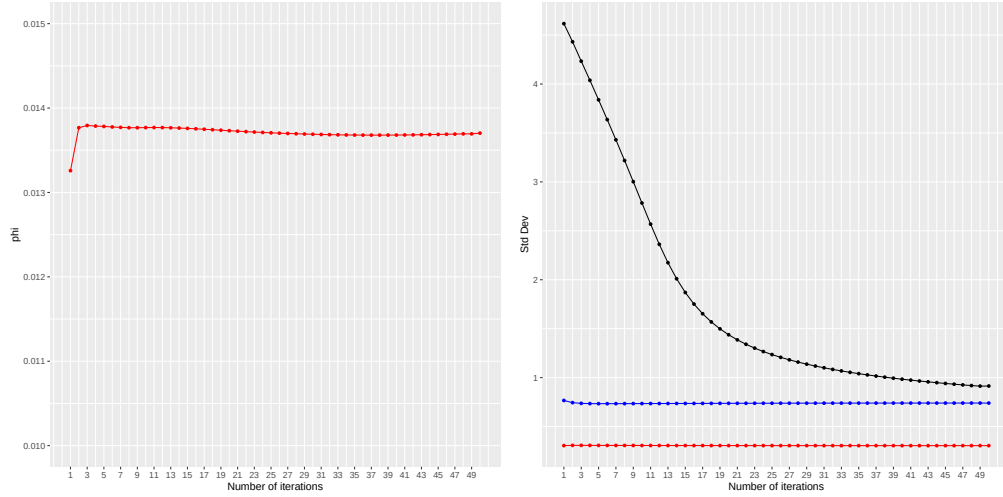


Figure 4.4: Estimation of the dispersion parameter  $\phi_2$  along algorithm iterations (left). Estimations of the standard deviations  $\sigma_\Delta$  (in black),  $\sigma_{\Gamma,1}$  (in blue) and  $\sigma_{\Gamma,2}$  (in red) along algorithm iterations (right).

Figures 4.5 to 4.8 depict the a posteriori corrections  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  together with the factor  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  for  $\mathbf{x} = (0,0)$ ,  $\mathbf{x} = (1,0)$ ,  $\mathbf{x} = (0,1)$  and  $\mathbf{x} = (1,1)$ , respectively. Horizontal lines correspond to a posteriori corrections based on  $\mathcal{H}_{i,T_i}^{\text{claim}}$  whereas trending curves illustrate the impact of incorporating past signal values in premium corrections.

For each value of  $\mathbf{x}$ , one sees that both a posteriori corrections  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  increase with the total number of claims  $N_{i1} + N_{i2} + N_{i3}$  observed over the last three periods. Furthermore,  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  also increases with the total number of dangerous driving events  $S_{i11} + S_{i12} + S_{i13}$  observed over the last three periods while it decreases with the normalized average speeds  $(S_{i21}, S_{i22}, S_{i23})$ . For instance, a policyholder belonging to the reference risk class  $\mathbf{x} = (0,0)$  who made no claim over the last three periods (here the last 9 months) and recording 10 dangerous driving events in total can see its expected claim frequency decreases by approximately 5% by using telematics data (i.e.  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}] \approx 95\%$ ) when its normalized average speeds are equal to 0.7 over the last three periods.

Finally, one can notice that the factor  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  is stable from one risk class to another. As expected, when comparing two drivers in two different risk classes, the one with the lower a priori claim frequency has higher a posteriori corrections  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$ .

Figures 4.5-4.8 show that despite relatively low estimated correlations, past signal values induce further segmentation among insured drivers, beyond that resulting from the inclusion of  $\mathcal{H}_{i,T_i}^{\text{claim}}$  in premium calculation. This is particularly encouraging for practical applications since with only two signals that can easily be communicated to policyholders, the credibility model is able to capture quite a large part of residual heterogeneity after just three periods when supplementing  $\mathcal{H}_{i,T_i}^{\text{claim}}$  with  $\mathcal{H}_{i,T_i}^{\text{signals}}$ .

## 5 Conclusion

Pay-How-You-Drive (PHYD) motor insurance complements traditional actuarial models with data reflecting driving behavior (e.g., vehicle average speed) and external risk factors (e.g., time of the day). Such data is recorded and transmitted by telematics devices embedded in each vehicle.

In this paper, we exploit the a posteriori nature of telematics data and their heterogeneity among insured drivers. We study a multivariate mixed model which extends the model describing the joint dynamics of telematics data and claim frequencies proposed by Denuit et al. (2019a) by replacing the estimation procedure with the ECM algorithm. The proposed model enables to combine telematics signals with various formats and claim counts, refining risk evaluations based on drivers' recent history.

Assessing our model through a case study with two telematics signals of different formats, we show that the proposed approach is already able to capture quite a large part of residual heterogeneity after just few months monitoring, allowing the pricing structure to become more reactive.

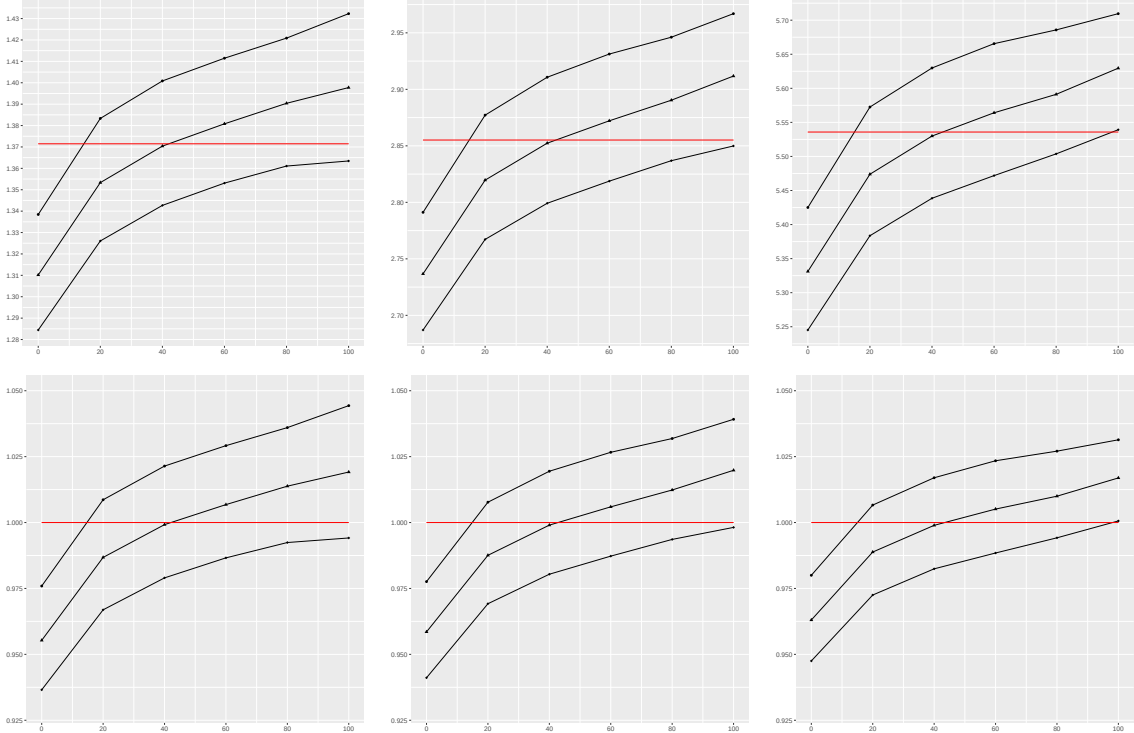


Figure 4.5:  $\mathbf{x} = (0, 0)$ . Left:  $N_{i1} + N_{i2} + N_{i3} = 0$ . Middle:  $N_{i1} + N_{i2} + N_{i3} = 1$ , Right:  $N_{i1} + N_{i2} + N_{i3} = 2$ . Top:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  (in black) and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  (in red) for  $(S_{i21} = 0.20, S_{i22} = 0.20, S_{i23} = 0.20)$  (circles),  $(S_{i21} = 0.40, S_{i22} = 0.40, S_{i23} = 0.40)$  (triangles),  $(S_{i21} = 0.70, S_{i22} = 0.70, S_{i23} = 0.70)$  (diamonds) and  $S_{i11} + S_{i12} + S_{i13} = 0, 20, 40, 60, 80, 100$ . Bottom:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$ .

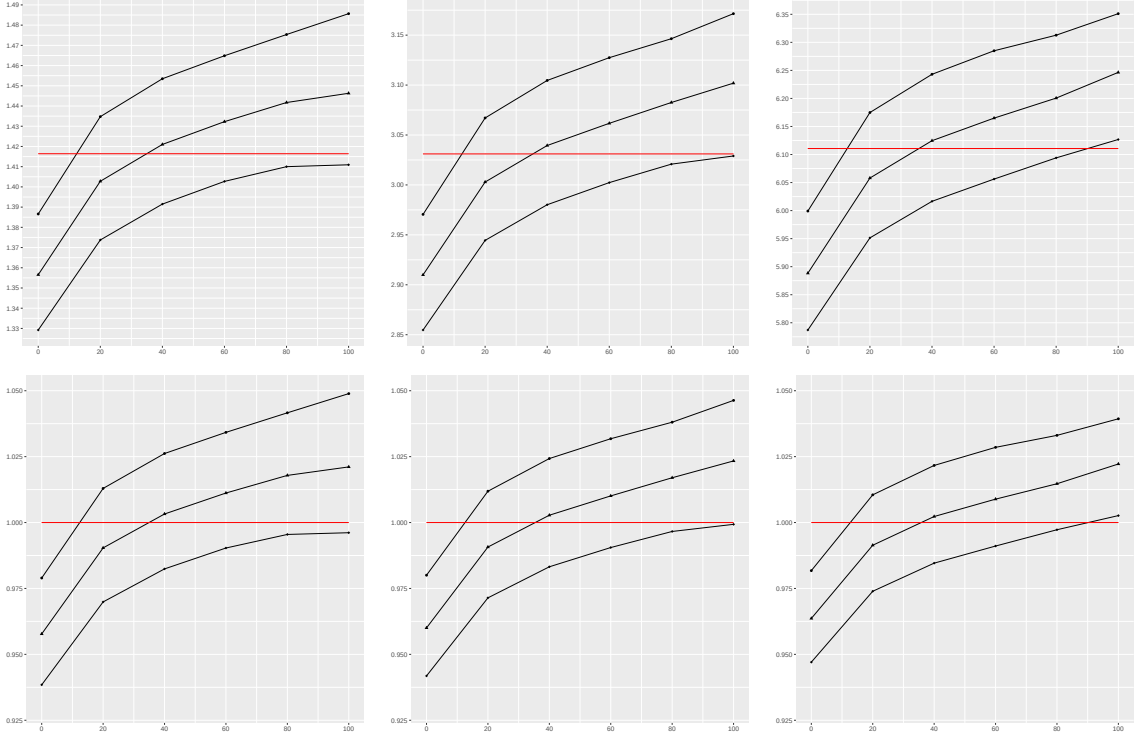


Figure 4.6:  $\mathbf{x} = (1, 0)$ . Left:  $N_{i1} + N_{i2} + N_{i3} = 0$ . Middle:  $N_{i1} + N_{i2} + N_{i3} = 1$ , Right:  $N_{i1} + N_{i2} + N_{i3} = 2$ . Top:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  (in black) and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  (in red) for  $(S_{i21} = 0.20, S_{i22} = 0.20, S_{i23} = 0.20)$  (circles),  $(S_{i21} = 0.40, S_{i22} = 0.40, S_{i23} = 0.40)$  (triangles),  $(S_{i21} = 0.70, S_{i22} = 0.70, S_{i23} = 0.70)$  (diamonds) and  $S_{i11} + S_{i12} + S_{i13} = 0, 20, 40, 60, 80, 100$ . Bottom:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$ .

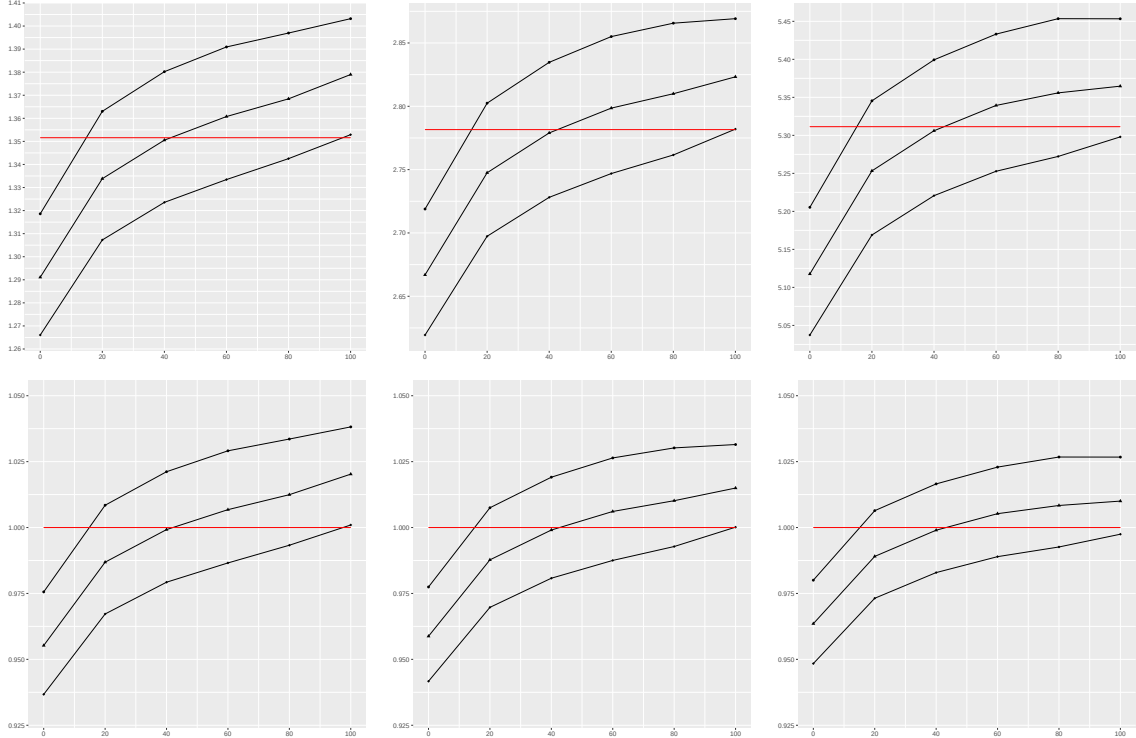


Figure 4.7:  $\mathbf{x} = (0, 1)$ . Left:  $N_{i1} + N_{i2} + N_{i3} = 0$ . Middle:  $N_{i1} + N_{i2} + N_{i3} = 1$ , Right:  $N_{i1} + N_{i2} + N_{i3} = 2$ . Top:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  (in black) and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  (in red) for  $(S_{i21} = 0.20, S_{i22} = 0.20, S_{i23} = 0.20)$  (circles),  $(S_{i21} = 0.40, S_{i22} = 0.40, S_{i23} = 0.40)$  (triangles),  $(S_{i21} = 0.70, S_{i22} = 0.70, S_{i23} = 0.70)$  (diamonds) and  $S_{i11} + S_{i12} + S_{i13} = 0, 20, 40, 60, 80, 100$ . Bottom:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$ .

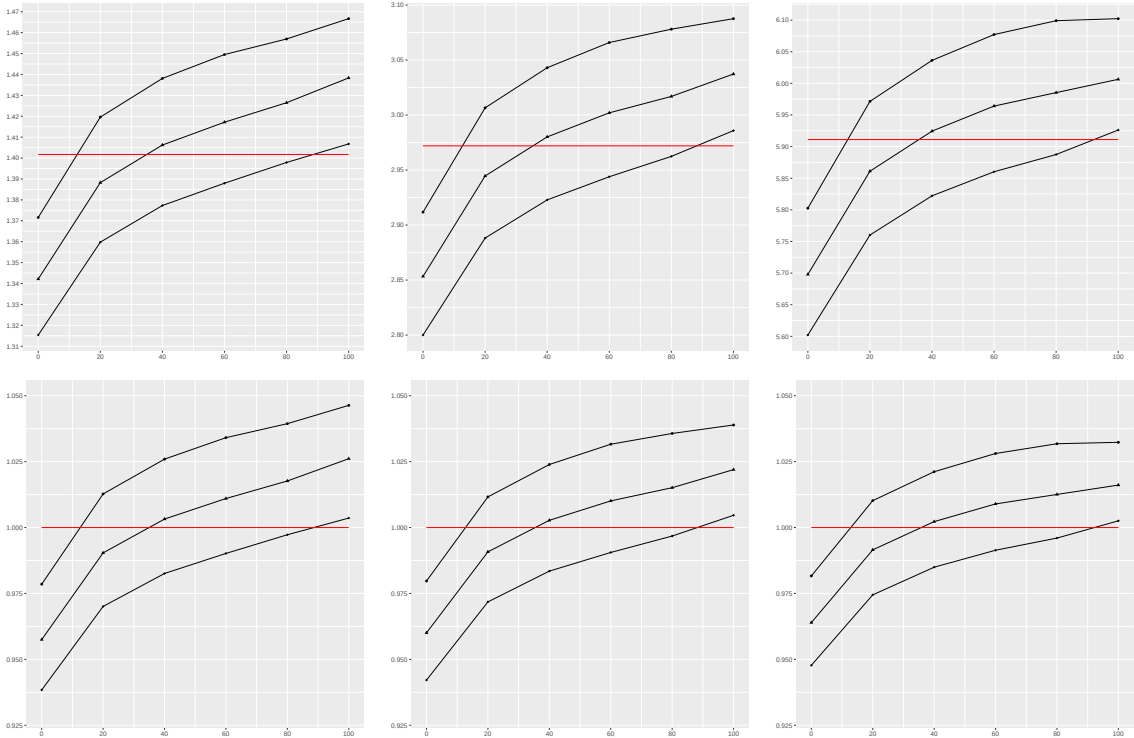


Figure 4.8:  $\mathbf{x} = (1, 1)$ . Left:  $N_{i1} + N_{i2} + N_{i3} = 0$ . Middle:  $N_{i1} + N_{i2} + N_{i3} = 1$ , Right:  $N_{i1} + N_{i2} + N_{i3} = 2$ . Top:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]$  (in black) and  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$  (in red) for  $(S_{i21} = 0.20, S_{i22} = 0.20, S_{i23} = 0.20)$  (circles),  $(S_{i21} = 0.40, S_{i22} = 0.40, S_{i23} = 0.40)$  (triangles),  $(S_{i21} = 0.70, S_{i22} = 0.70, S_{i23} = 0.70)$  (diamonds) and  $S_{i11} + S_{i12} + S_{i13} = 0, 20, 40, 60, 80, 100$ . Bottom:  $E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}]/E[\exp(\Delta_i)|\mathcal{H}_{i,T_i}^{\text{claim}}]$ .

## References

- Ayuso, M., Guillen, M., Perez-Marin, A. M. (2016). Telematics and gender discrimination: Some usage-based evidence on whether men's risk of accidents differs from women's. *Risks* 4, 1-10.
- Ayuso, M., Guillen, M., Nielsen, J.P. (2019). Improving automobile insurance ratemaking using telematics: Incorporating mileage and driver behaviour data. *Transportation* 46, 735-752.
- Bolderdijk, J., Knockaert, J., Steg, L., Verhoef, E. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. *Accident: Analysis and Prevention* 43, 1181-1186.
- Boucher, J.-P., Denuit, M., Guillen, M. (2007). Risk classification for claim counts: A comparative analysis of various zero-inflated Mixed Poisson and Hurdle models. *North American Actuarial Journal* 11, 110-131.
- Boucher, J.-Ph., Denuit, M. (2008). Credibility premiums for the zero-inflated Poisson model and new hunger for bonus interpretation. *Insurance: Mathematics and Economics* 42, 727-735.
- Boucher, J. P., Perez-Marin, A. M., Santolino, M. (2013). Pay-as-you-drive insurance: The effect of the kilometers on the risk of accident. *Anales del Instituto de Actuarios Españoles* 19, 135-154.
- Dempster, A. P., N. M. Laird, D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society – Series B: Methodological* 39, 1–22.
- Denuit, M., Guillen, M., Trufin, J. (2019a). Multivariate credibility modeling for usage-based motor insurance pricing with behavioral data. *Annals of Actuarial Science* 13, 378-399.
- Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Wiley, New York.
- Denuit, M., Hainaut, D., Trufin, J. (2019b). *Effective Statistical Learning Methods for Actuaries Volume 1: GLM and Extensions* Springer Actuarial Lecture Notes Series.
- Fung, T.C., Badescu, A.L., Lin, X.S. (2019). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin* 49, 647-688.
- Fung, T.C., Badescu, A.L., Lin, X.S. (2020). A new class of severity regression models with an application to IBNR prediction. *North American Actuarial Journal*, in press.

- Gao, G., Meng, S., Wuthrich, M. V. (2019a). Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal* 2019, 143-162.
- Gao, G., Wuthrich, M. V., Yang, H. (2019b). Evaluation of driving risk at different speeds. *Insurance: Mathematics and Economics* 88, 108-119.
- Guillen, M., Nielsen, J.P., Ayuso, M., Perez-Marin, A.M. (2019). The use of telematics devices to improve automobile insurance rates. *Risk Analysis* 39, 662-672.
- Guillen, M., Nielsen, J. P., Perez-Marin, A. M., Elpidorou, V. (2020). Can automobile insurance telematics predict the risk of near-miss events? *North American Actuarial Journal* 24, 141-152.
- Grari, V., Lamprier, S., Detyniecki, M. (2020). Fairness-Aware Neural Rényi Minimization for Continuous Features. *Proceedings of the 29th International Joint Conference on Artificial Intelligence IJCAI-20*, 2262-2268.
- Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 80, 267-278.
- Ortiz, F.M., Sammarco, M., Costa, L.H.M.K., Detyniecki, M. (2020). Vehicle telematics via exteroceptive sensors: A survey. Available from <https://arxiv.org/abs/2008.12632>
- Pechon, F., Trufin, J., Denuit, M. (2018). Multivariate modelling of household claim frequencies in motor third-party liability insurance. *ASTIN Bulletin* 48, 969-993.
- Pechon, F., Denuit, M., Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal* 9, 575-602.
- Pechon, F., Denuit, M., Trufin, J. (2020). Home and Motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, in press.
- Tselentis, D. I., Yannis, G., Vlahogianni, E. I. (2016). Innovative insurance schemes: Pay as/how you drive. *Transportation Research Procedia* 14, 362-371.
- Varadhan, R., C. Roland (2008). Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scandinavian Journal of Statistics* 35, 335-353.
- Verbelen, R., Antonio, K., Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67 1275-1304.
- Wahlstrom, J., Skog, I., Handel, P. (2015). Driving behavior analysis for smartphone-based insurance telematics. *Proceedings of the 2nd Workshop on Physical Analytics WPA '15*, 19-24.
- Weidner, W., Transchel, F. W. G., Weidner, R. (2016). Classification of scale-sensitive telematic observables for risk individual pricing. *European Actuarial Journal* 6, 3-24.

- Weidner, W., Transchel, F. W., Weidner, R. (2017). Telematic driving profile classification in car insurance pricing. *Annals of Actuarial Science* 11, 213-236.
- Wuthrich, M. V. (2017). Covariate selection from telematics car driving data. *European Actuarial Journal* 7, 89-108.

# APPENDIX

## A Implementation of the ECM algorithm

Denote as  $\mathcal{O}_i$  all observations for policyholder  $i$ , that is,

$$\mathcal{O}_i = \{n_{i1}, \dots, n_{iT_i}, s_{i11}, \dots, s_{i1T_i}, \dots, s_{iq1}, \dots, s_{iqT_i}\}.$$

Also, let  $\mathcal{S}_i$  gather all the scores for policyholder  $i$ , that is,

$$\mathcal{S}_i = \{\eta_{i1}, \dots, \eta_{iT_i}, \nu_{i11}, \dots, \nu_{i1T_i}, \dots, \nu_{iq1}, \dots, \nu_{iqT_i}\}.$$

To fix the ideas, we assume that these scores have been produced by a GLMM. The procedure easily extends to other supervised learning models such as GAMs or boosting methods, for instance.

The ECM algorithm then proceeds as follows:

### Initialize:

- Run a GLMM with an intercept, only, on each response and signal, separately, so that we get

(1) Scores  $\eta_{it}^{(0)}$  and  $\nu_{ikt}^{(0)}$  gathered in  $\mathcal{S}_i^{(0)}$ ;

(2) A diagonal variance-covariance matrix  $\widehat{\Sigma}^{(0)}$  (with diagonal elements coming from marginal GLMM analyses).

- Compute the expectation of each response and signal given the observed values  $\mathcal{O}_i$  and scores  $\mathcal{S}_i^{(0)}$  using  $\widehat{\Sigma}^{(0)}$ , that is, compute  $\widehat{\mathbb{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$ ,  $\widehat{\mathbb{E}}[\Gamma_{i1} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$ ,  $\dots$ ,  $\widehat{\mathbb{E}}[\Gamma_{iq} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$ . The calculation of these conditional expectations is explained in Appendix B below.
- Fit GLMs with  $\widehat{\mathbb{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$ ,  $\widehat{\mathbb{E}}[\Gamma_{i1} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$ ,  $\dots$ ,  $\widehat{\mathbb{E}}[\Gamma_{iq} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]$  as offsets, separately. This produces new scores  $\eta_{it}^{(1)}$  and  $\nu_{ikt}^{(1)}$  gathered in  $\mathcal{S}_i^{(1)}$ .
- Estimate the variance-covariance matrix  $\widehat{\Sigma}^{(1)}$  whose diagonal elements are

$$\begin{aligned} \widehat{\text{Var}}[\Delta_i] &= \widehat{\mathbb{E}}[\widehat{\text{Var}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] + \widehat{\text{Var}}[\widehat{\mathbb{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] \\ \widehat{\text{Var}}[\Gamma_{ik}] &= \widehat{\mathbb{E}}[\widehat{\text{Var}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] + \widehat{\text{Var}}[\widehat{\mathbb{E}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]], \quad k = 1, \dots, q \end{aligned}$$

and out-of-diagonal elements are

$$\begin{aligned} \widehat{\text{Cov}}[\Delta_i, \Gamma_{ik}] &= \widehat{\mathbb{E}}[\widehat{\text{Cov}}[\Delta_i, \Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] + \widehat{\text{Cov}}[\widehat{\mathbb{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}], \widehat{\mathbb{E}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] \\ &\quad k = 1, \dots, q \\ \widehat{\text{Cov}}[\Gamma_{ik_1}, \Gamma_{ik_2}] &= \widehat{\mathbb{E}}[\widehat{\text{Cov}}[\Gamma_{ik_1}, \Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] + \widehat{\text{Cov}}[\widehat{\mathbb{E}}[\Gamma_{ik_1} | \mathcal{O}_i, \mathcal{S}_i^{(0)}], \widehat{\mathbb{E}}[\Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(0)}]] \\ &\quad k_1 \neq k_2 \in \{1, \dots, q\}. \end{aligned}$$

We then use the fact that

$$\begin{aligned}
\widehat{\text{Var}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}] &= \widehat{\text{E}}[\Delta_i^2 | \mathcal{O}_i, \mathcal{S}_i^{(0)}] - \widehat{\text{E}}^2[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \\
\widehat{\text{Var}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] &= \widehat{\text{E}}[\Gamma_{ik}^2 | \mathcal{O}_i, \mathcal{S}_i^{(0)}] - \widehat{\text{E}}^2[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \\
&\quad j = 1, \dots, q \\
\widehat{\text{Cov}}[\Delta_i, \Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] &= \widehat{\text{E}}[\Delta_i \Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] - \widehat{\text{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \widehat{\text{E}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \\
&\quad k = 1, \dots, q \\
\widehat{\text{Cov}}[\Gamma_{ik_1}, \Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] &= \widehat{\text{E}}[\Gamma_{ik_1} \Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] - \widehat{\text{E}}[\Gamma_{ik_1} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \widehat{\text{E}}[\Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(0)}] \\
&\quad k_1 \neq k_2 \in \{1, \dots, q\}.
\end{aligned}$$

The conditional expectations appearing in the variances and covariances can be computed as explained in Appendix B, with a slight adaptation of the integrand. The estimated  $\sigma_{\Delta}^2$ ,  $\sigma_{\Gamma,k}^2$ ,  $\sigma_{\Delta,\Gamma,k}$  and  $\sigma_{\Gamma,k_1,k_2}$  then follow by using sample means, variances and covariances since sample moments correspond to the maximum-likelihood estimates in the Normal case.

**Cycle:** as long as the chosen stopping criterion is not satisfied, iterate

- E-Step: Compute the expectation of each response and signal given the observed values  $\mathcal{O}_i$  and scores  $\mathcal{S}_i^{(l)}$  using  $\widehat{\Sigma}^{(l)}$ , that is, compute  $\widehat{\text{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$ ,  $\widehat{\text{E}}[\Gamma_{i1} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$ ,  $\dots$ ,  $\widehat{\text{E}}[\Gamma_{iq} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$ .
- CM1-Step: Fit GLMs with  $\widehat{\text{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$ ,  $\widehat{\text{E}}[\Gamma_{i1} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$ ,  $\dots$ ,  $\widehat{\text{E}}[\Gamma_{iq} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]$  as offsets, separately, to obtain new scores  $\eta_{it}^{(l+1)}$  and  $\nu_{ikt}^{(l+1)}$  gathered in  $\mathcal{S}_i^{(l+1)}$ .
- CM2-Step: Estimate the variance-covariance matrix  $\Sigma$  by  $\widehat{\Sigma}^{(l+1)}$  whose diagonal elements are

$$\begin{aligned}
\widehat{\text{Var}}[\Delta_i] &= \widehat{\text{E}}[\widehat{\text{Var}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] + \widehat{\text{Var}}[\widehat{\text{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] \\
\widehat{\text{Var}}[\Gamma_{ik}] &= \widehat{\text{E}}[\widehat{\text{Var}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] + \widehat{\text{Var}}[\widehat{\text{E}}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]], \quad k = 1, \dots, q
\end{aligned}$$

and out-of-diagonal elements are

$$\begin{aligned}
\widehat{\text{Cov}}[\Delta_i, \Gamma_{ikj}] &= \widehat{\text{E}}[\widehat{\text{Cov}}[\Delta_i, \Gamma_{ikj} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] + \widehat{\text{Cov}}[\widehat{\text{E}}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i^{(l)}], \widehat{\text{E}}[\Gamma_{ikj} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] \\
&\quad j = 1, \dots, q \\
\widehat{\text{Cov}}[\Gamma_{ik_1}, \Gamma_{ik_2}] &= \widehat{\text{E}}[\widehat{\text{Cov}}[\Gamma_{ik_1}, \Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] + \widehat{\text{Cov}}[\widehat{\text{E}}[\Gamma_{ik_1} | \mathcal{O}_i, \mathcal{S}_i^{(l)}], \widehat{\text{E}}[\Gamma_{ik_2} | \mathcal{O}_i, \mathcal{S}_i^{(l)}]] \\
&\quad k_1 \neq k_2 \in \{1, \dots, q\}.
\end{aligned}$$

## B Evaluation of the conditional expectations involved in the ECM algorithm

The ECM approach involves conditional expectations of the random effects in each E-step. Here, we carefully explain how to compute  $\text{E}[\Delta_i | \mathcal{O}_i, \mathcal{S}_i]$  and  $\text{E}[\Gamma_{ik} | \mathcal{O}_i, \mathcal{S}_i]$  so that the reader

can easily adapt the formulas to the other conditional expectations appearing in the ECM algorithm.

We have

$$\mathbb{E} [\Delta_i | \mathcal{O}_i, \mathcal{S}_i] = \int_{-\infty}^{\infty} \delta f_{\Delta_i | \mathcal{O}_i, \mathcal{S}_i}(\delta) d\delta$$

where  $f_{\Delta_i | \mathcal{O}_i, \mathcal{S}_i}$  denotes the conditional probability density function of  $\Delta_i$  given past observations  $\mathcal{O}_i$  and past values of the scores  $\mathcal{S}_i$ . Assume that the signals  $S_{ikt}$  are continuous, with conditional probability density function  $f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(\cdot)$  given  $\Gamma_{ik} = \gamma_k$ . For discrete signals, the conditional probability density function is replaced with the conditional probability mass function. The conditional expectation can then be expressed as the ratio  $\mathbb{E} [\Delta_i | \mathcal{O}_i, \mathcal{S}_i] = \frac{\tilde{A}_{(\Delta)}}{\tilde{B}}$  with

$$\begin{aligned} \tilde{A}_{(\Delta)} = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \delta \prod_{t=1}^{T_i} \left( \mathbb{P}[N_{it} = n_{it} | \Delta_i = \delta] \prod_{k=1}^q f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(s_{ikt}) \right) \\ & f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q \end{aligned}$$

and

$$\begin{aligned} \tilde{B} = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \left( \mathbb{P}[N_{it} = n_{it} | \Delta_i = \delta] \prod_{k=1}^q f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(s_{ikt}) \right) \\ & f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q. \end{aligned}$$

Under the assumptions of our model, we know that given  $\Delta_i = \delta$ ,  $N_{i\bullet} = \sum_{t=1}^{T_i} N_{it}$  obeys the  $\text{Poisson}(\lambda_{i\bullet} \exp(\delta))$  distribution, where  $\lambda_{i\bullet} = \sum_{t=1}^{T_i} \lambda_{it}$ . Therefore,  $\mathbb{E} [\Delta_i | \mathcal{O}_i, \mathcal{S}_i] = \frac{\tilde{A}_{(\Delta)}}{\tilde{B}} = \frac{A_{(\Delta)}}{B}$  with

$$\begin{aligned} A_{(\Delta)} = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \delta (\exp(\delta))^{n_{i\bullet}} \exp(-\lambda_{i\bullet} \exp(\delta)) \prod_{t=1}^{T_i} \prod_{k=1}^q f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(s_{ikt}) \\ & f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q \end{aligned}$$

and

$$\begin{aligned} B = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\exp(\delta))^{n_{i\bullet}} \exp(-\lambda_{i\bullet} \exp(\delta)) \prod_{t=1}^{T_i} \prod_{k=1}^q f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(s_{ikt}) \\ & f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q \end{aligned}$$

where  $n_{i\bullet} = \sum_{t=1}^{T_i} n_{it}$  stands for the observed claim totals  $N_{i\bullet}$ .

Similarly, we can write  $\mathbb{E} [\Gamma_{ik_0} | \mathcal{O}_i, \mathcal{S}_i] = \frac{A_{(\Gamma_{k_0})}}{B}$  for each signal  $\Gamma_{ik_0}$  with

$$\begin{aligned} A_{(\Gamma_{k_0})} = & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \gamma_{k_0} (\exp(\delta))^{n_{i\bullet}} \exp(-\lambda_{i\bullet} \exp(\delta)) \prod_{t=1}^{T_i} \prod_{k=1}^q f_{S_{ikt} | \Gamma_{ik} = \gamma_k}(s_{ikt}) \\ & f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q \end{aligned}$$

and

$$B = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\exp(\delta))^{n_{i\bullet}} \exp(-\lambda_{i\bullet} \exp(\delta)) \prod_{t=1}^{T_i} \prod_{k=1}^q f_{S_{ikt}|\Gamma_{ik}=\gamma_k}(s_{ikt}) \\ f_{\Sigma}(\delta, \gamma_1, \dots, \gamma_q) d\delta d\gamma_1 \dots d\gamma_q.$$

Conditional expectations appearing at iteration  $l = 0, 1, 2, \dots$  of the ECM algorithm (iteration  $l = 0$  corresponding to the initialization) are then obtained by replacing  $\mathcal{S}_i$  with  $\mathcal{S}_i^{(l)}$ . The integrals involved in these formulas can be computed using quadratures, as explained in Pechon et al. (2018, 2019, 2020). Gauss-Hermite quadratures revealed itself to be fast, especially when used in conjunction with the package **Rcpp** which allows to write chunk of codes in **C++** inside an **R** script enabling a drastic reduction in computational time (up to 30 times faster).