MULTIVARIATE MODELLING OF HOUSEHOLD CLAIM FREQUENCIES IN MOTOR THIRD-PARTY LIABILITY INSURANCE

Julien Trufin (joint work with Michel Denuit and Florian Pechon)
Université Libre de Bruxelles

Chairs Days: Insurance, Actuarial Science, Data and Models

June 11 2018

Paris

Agenda

Julien Trufin (joint work with Michel Denuit and Florian Pechon)
Université Libre de Bruxelles

Outline

Julien Trufin (joint work with Michel Denuit and Florian Pechon)
Université Libre de Bruxelles

- Non-life insurance pricing is usually performed product by product:
  - Careful risk assessment for each product, **in isolation**
  - Letting expected claims frequencies vary according to **policyholder's risk profile**
- Typical explanatory variables are: Age of insured, age of car, gender, use of the car, location, split of premium, occupation, etc.
- Claim frequencies: typically modelled using the **GLM/GAM** framework

$$N_{it} \sim \mathcal{P}(\lambda_{it})$$

$$\log \lambda_{it} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \sum_j f_j(x_{ij}) + f(x_{lat,i}, x_{long,i})$$

- The **Poisson regression** model imposes **equidispersion** once the explanatory variables have been taken into account, that is, $\mathbb{E}[N_{it}|\boldsymbol{x_{it}}] = \mathbb{V}[N_{it}|\boldsymbol{x_{it}}]$.
- **Unobserved heterogeneity** will often cause the variance to exceed the mean (a phenomenon termed *overdispersion*).
- Introduction of a **random effect** Θ_i s.t. $\mathbb{E}[\Theta_i] = 1$:

$$\text{Given } \Theta_i = \theta_i,\ N_{it} \sim \mathcal{P}(\lambda_{it}\theta_i) \quad \text{independent}\,\forall i$$

$$\Longrightarrow \mathbb{V}(N_{it}) = \lambda_{it} + \lambda_{it}^2\mathbb{V}(\Theta_i) > \lambda_{it}$$

- **A posteriori claim frequency**

$$\mathrm{E}\left[\lambda_{it}\Theta_i|N_{i1}=n_1, N_{i2}=n_2, \ldots, N_{ik}=n_k\right] = \lambda_{it}\int_0^\infty \theta_i f_{\Theta_i}(\theta_i|N_{i1}=n_1, N_{i2}=n_2, \ldots, N_{ik}=n_k)d\theta_i$$

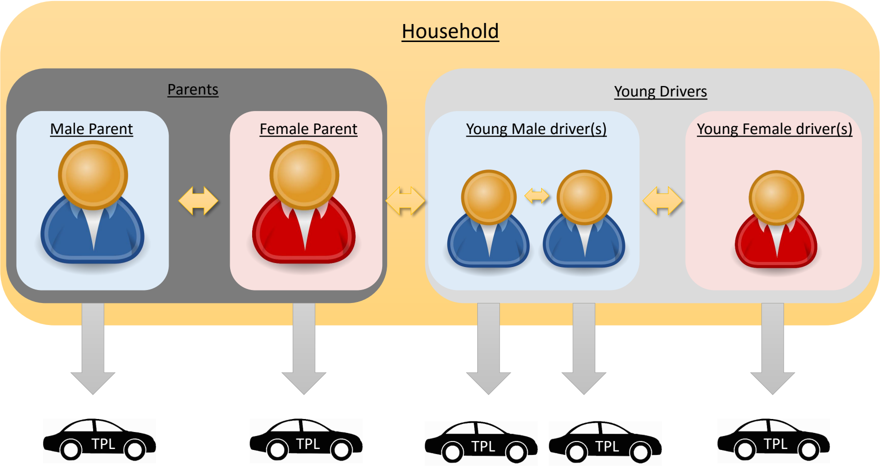
Julien Trufin (joint work with Michel Denuit and Florian Pechon)
Université Libre de Bruxelles

- For a given product, ratemaking is usually considered on a **policyholder level**.
- Idea: working at the **household level**.
- Let us now identify (subject to the limitations of the data basis)
  - The **kind of profile** of the policyholder.
  - Two main populations: **parents** and **young drivers**.
  - Which policyholders belong to **the same household**.

$\Longrightarrow$ Joint modelling on household level, taking into account **correlations** between numbers of claims of **policyholders from the same household**.

- For a given product, ratemaking is usually considered on a **policyholder level**.
- Idea: working at the **household level**.
- Let us now identify (subject to the limitations of the data basis)
  - The **kind of profile** of the policyholder.
  - Two main populations: **parents** and **young drivers**.
  - Which policyholders belong to **the same household**.

$\Longrightarrow$ Joint modelling on household level, taking into account **correlations** between numbers of claims of **policyholders from the same household**.



- Only informations about individuals covered by AXA is available.
- Let us look at the **age of adults living with young drivers** that have a TPL policy. This will give an idea of the age at which one can expect young drivers in the household with an AXA insurance policy.

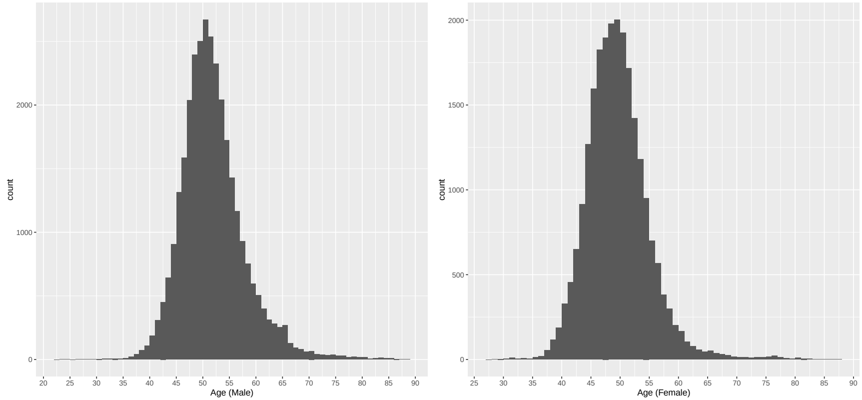


Figure: Left (Right): Age of male (female) parent living in a household with a policyholder younger than ≤ 23 years.

- Choosing as cutoff points of 38 and 56 allows to have a homogeneous group of adults that could possibly have young driver with an insurance policy.

Modelling TPL on a household level

Higher claim frequency for younger drivers

- Let us find the age at which a child has similar claim frequency than adults.
- This will help determine the break off point for population P3 and P4.

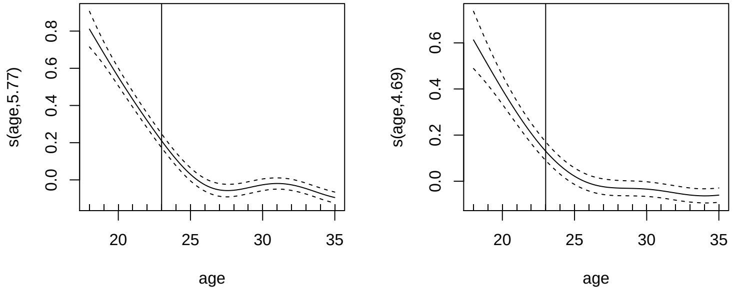
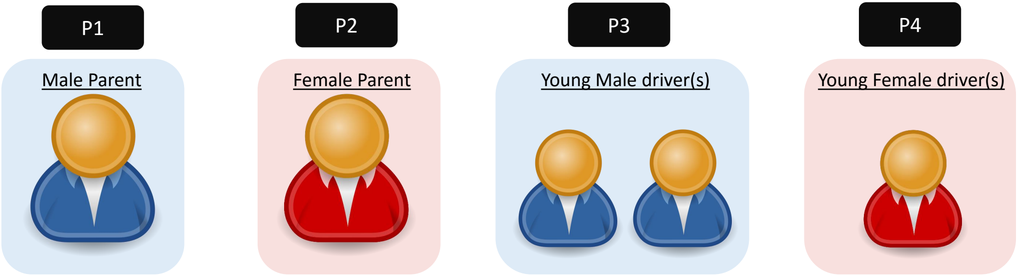


Figure: Impact of age on estimated claim frequency. Left: Male. Right: Female

- Choosing a break off point of 23 years for populations P3 and P4 will help modelling the high risk profiles.

- We wish to do a separate analysis of these **4 subpopulations**:
  - P1 Male parents (or who could have children, given their age).
  - P2 Female parents (or who could have children, given their age).
  - P3 Males younger than 23 years and living with their parents.
  - P4 Females younger than 23 years and living with their parents.



- The analysis can be divided into two parts:
  1. Estimate **a priori claim frequencies** based on
     - Age of policyholder;
     - Power of car;
     - Usage of car;
     - Split of premium;
     - ZIP code of residence;
     - ...
     - **Household characteristics**.

     $\Longrightarrow$ Observable characteristics. However, some *unobservable characteristics* also differentiate the policyholders.
  2. Adjust the a priori claim frequencies with **a posteriori corrections** based on the **past claims history** which reveals these *unobservable characteristics*.
     - Past claims history of the policyholder...
     - ...and **of other policyholders from the household**.

Outline

- Assume that the number of claims in TPL can be modelled using a Poisson regression and GAM (using splines for non-linear effects) by

$$\log \mathbb{E}[N] = \log Expo + \beta_{usage\_profes} x_{usage\_profes} + \boldsymbol{\beta}_{split} \boldsymbol{x}_{split} + \boldsymbol{\beta}_{litigation} \boldsymbol{x}_{litigation} + s(age, by = gender) + s(power) + s(Latitude, Longitude)$$

- This will be fitted on the four populations separately.
- Due to smaller population size, the spatial effect is fitted separately on the whole portfolio, and used later as offset in the four fits.

MULTIVARIATE MODELLING OF HOUSEHOLD CLAIM FREQUENCIES IN MOTOR THIRD-PARTY LIABILITY INSURANCE

Julien Trufin (joint work with Michel Denuit and Florian Pechon) Université Libre de Bruxelles

Introduction

A priori claim frequency

A posteriori claim frequency

Applications

Variable	P1	P2
Intercept	-0.15118	-0.15180
	0.01016	*0.01209*
usage_profes	0.16587	0.13630
	0.02376	*0.03235*
splitM	0.18378	0.23251
	0.02110	*0.02341*
splitS	0.16757	0.15726
	0.01595	*0.01984*
splitQ	0.32511	0.34930
	0.01832	*0.02171*
litigation2	0.26796	0.26050
	0.04011	*0.04918*
litigation4	0.63416	0.43684
	0.05301	*0.08417*

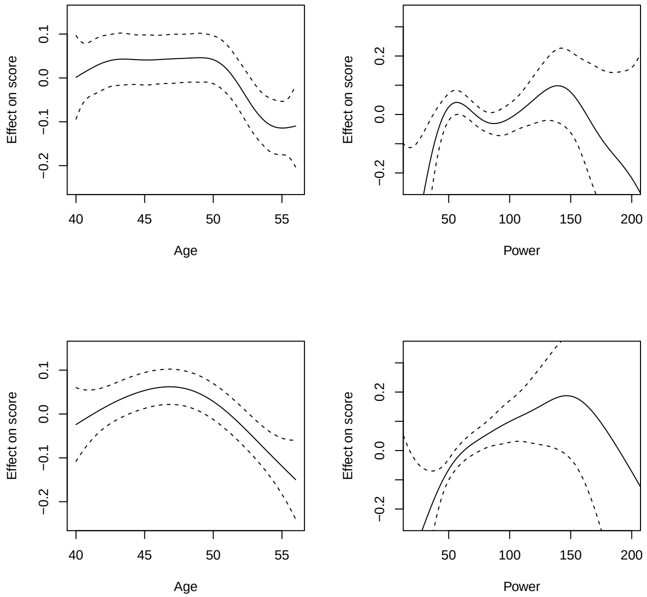
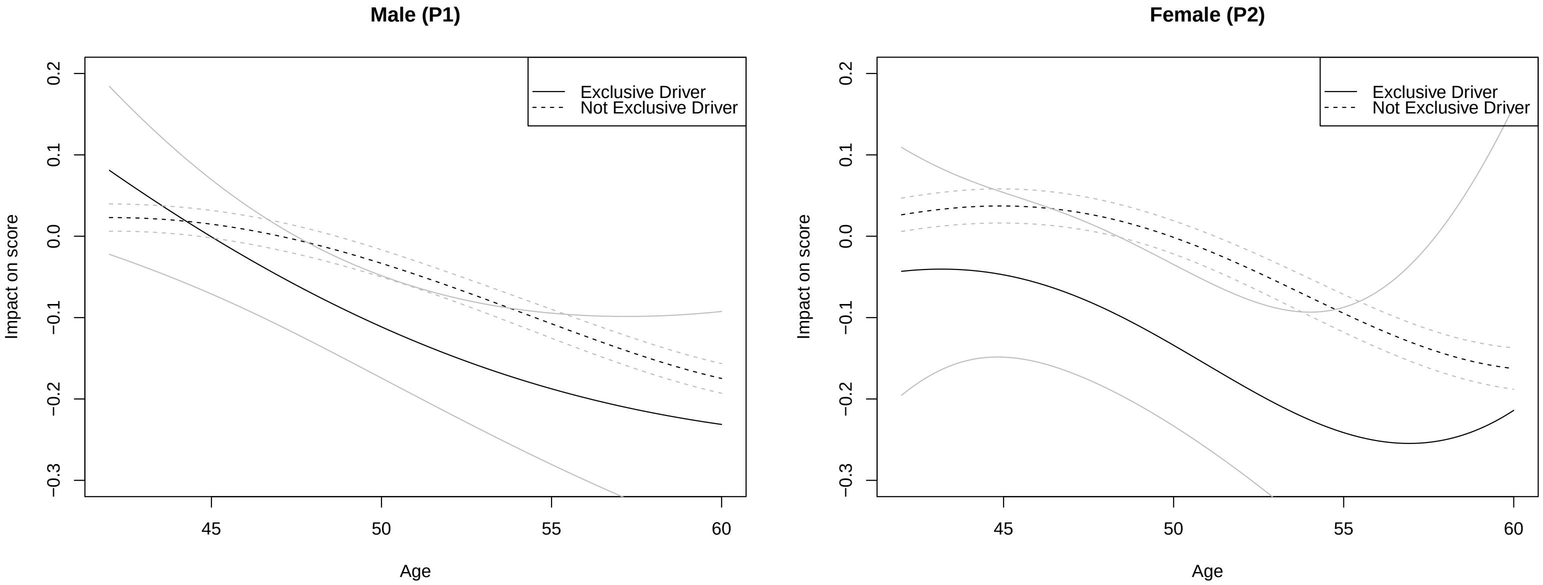


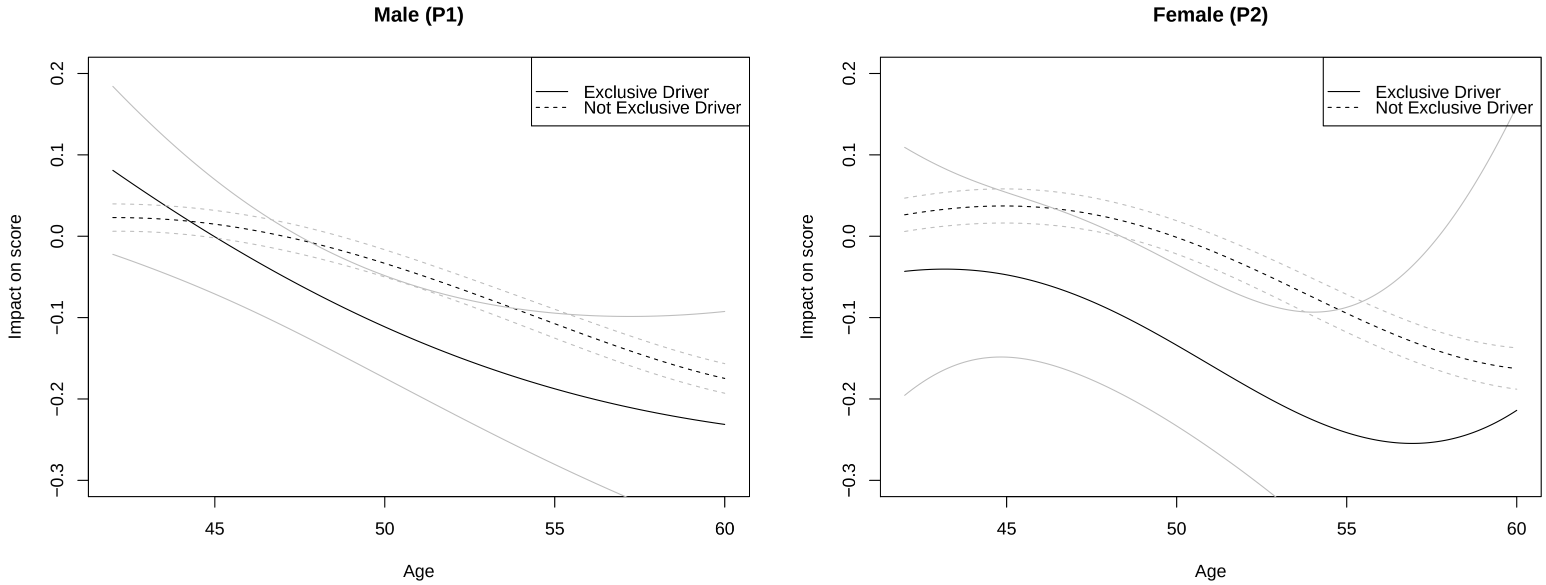
Figure: Continuous covariates (top: P1, bottom: P2)

- Let us analyse the impact of the **age of policyholder** on the claim frequency. Let us distinguish policyholder **with and without the exclusive driver clause.**[1]



[1] The exclusive driver clause enforces that only the policyholder and his/her spouse are allowed to drive with the insured car.

- Let us analyse the impact of the **age of policyholder** on the claim frequency. Let us distinguish policyholder **with and without the exclusive driver clause.**[1]



$\Longrightarrow$ **The “bump"/hill around 45-50 years does not appear for exclusive drivers.**

$\Longrightarrow$ **This suggests that these higher claim frequencies come from other people using the car, possibly young drivers.**

[1] The exclusive driver clause enforces that only the policyholder and his/her spouse are allowed to drive with the insured car.

- We wish to add a new variable in the a priori modelling which takes into account
  - the number of insureds inside a household,
  - whether there are young drivers with their own AXA policy in the household and
  - if the parents have agreed on a exclusive driver clause (which allows only the insured and its spouse to drive the car, excluding the young drivers).

Let us use the information about the number of policyholders in the household and whether the policy has a "exclusive driver" clause.

For **male and female parents** we found:

	Estimate	Std. Error	z value	Pr(>|z|)
1 Parent No Y.D.[2] non exclusive (Intercept)	0.0197	0.0055	3.56	0.0004
1 Parent No Y.D. exclusive OR 1 Parent with 1+ Y.D.	-0.0597	0.0290	-2.06	0.0395
2 Parents No Y.D. exclusive OR 2 Parents with 1+ Y.D.	-0.1380	0.0555	-2.49	0.0129
2 Parents No Y.D. non exclusive	-0.0937	0.0151	-6.19	0.0000

$\Longrightarrow$ **Household with several policyholders all covered by AXA are less risky.**

In particular in households with young drivers holding a policy.

$\leadsto$**Legal or material reason that prevents young drivers to drive with their parents' car**

[2] Y.D. = young driver

- Let us see the type of risk profiles that are found inside a same household (a priori claim frequencies).
- We consider **three kinds of risk profiles** (low, medium, high), based on quantiles of the estimated a priori claim frequencies.

Risk Profile of Husband	Risk Profile of Wife	Low	Medium	High
	Low	**23.69%**	8.28%	1.71%
	Medium	8.34%	**17.15%**	8.28%
	High	1.55%	8.29%	**22.71%**

Table: Relative exposure by risk profile of male (rows) and female (columns). Computed using only households with exactly one male (P1) and one female (P2), regardless if there are any young drivers.

$\Longrightarrow$ **Couples tend to have a similar risk profile** (obvious for age or place of residence, but also true in general.)

- Hence, we cannot rely on correlation coefficients for claim counts.

Outline

- Let us first consider a simpler framework, namely all the households composed by one male parent and/or one female parent [3].
- In the following, we use a multivariate Poisson mixture.
- This model is based on the following assumptions (where T denotes the number of observation periods):
  1. For $j \in \{1,2\}$, given $\Theta_h^{Pj} = \theta$, the random variables $N_{h1}^{Pj}, N_{h2}^{Pj}, \ldots, N_{hT}^{Pj}$ are independent, Poisson distributed with respective means $\lambda_{h1}^{Pj}\theta, \lambda_{h2}^{Pj}\theta, \ldots, \lambda_{hT}^{Pj}\theta$.
  2. Given $(\Theta_h^{P1}, \Theta_h^{P2})$, the random variables $N_{h1}^{P1}, N_{h2}^{P1}, \ldots, N_{hT}^{P1}$ and $N_{h1}^{P2}, N_{h2}^{P2}, \ldots, N_{hT}^{P2}$ are independent.
  3. The pairs $(\Theta_h^{P1}, \Theta_h^{P2})$ are independent and identically distributed, with common joint probability density function $f_{\boldsymbol{\Theta}}$, $\mathrm{E}[\Theta_h^{Pj}] = 1$ for $j \in \{1,2\}$ and variance-covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\Theta}} = \begin{pmatrix} (\sigma_{\boldsymbol{\Theta}}^{P1})^2 & \sigma_{\boldsymbol{\Theta}}^{P:P} \\ \sigma_{\boldsymbol{\Theta}}^{P:P} & (\sigma_{\boldsymbol{\Theta}}^{P2})^2 \end{pmatrix}.$$

- We also use the correlation coefficient $\rho_{\boldsymbol{\Theta}}^{P:P} = \frac{\sigma_{\boldsymbol{\Theta}}^{P:P}}{\sigma_{\boldsymbol{\Theta}}^{P1}\sigma_{\boldsymbol{\Theta}}^{P2}}$ in addition to the covariance $\sigma_{\boldsymbol{\Theta}}^{P:P}$.

[3] There may be young drivers, but we do not consider them here.

Julien Trufin (joint work with Michel Denuit and Florian Pechon)
Université Libre de Bruxelles

- Let $\mathcal{H}_1$ (resp. $\mathcal{H}_2$) denote the set of all households comprising a member in P1 (resp. P2), i.e. with husband/father (resp. mother/wife) insured.
- Then, $\mathcal{H}_{12} = \mathcal{H}_1 \cap \mathcal{H}_2$ corresponds to the set of all households with both husband and wife insured.
- In addition, define the set $\mathcal{H}_{1\backslash 2} = \mathcal{H}_1 \setminus \mathcal{H}_2$ of all households with husband insured, but not his wife, and the set $\mathcal{H}_{2\backslash 1} = \mathcal{H}_2 \setminus \mathcal{H}_1$ of all households with wife insured, but not her husband.

We have compared two (bivariate) parametric models:

- **Bivariate Poisson-LogNormal model**
- Bivariate Poisson-Gamma model

The likelihood can be written as

$$\mathcal{L}(\boldsymbol{\Sigma}) = \mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{L}_3$$

where

$$\begin{aligned}
\mathcal{L}_1 &= \prod_{h \in \mathcal{H}_{12}} \mathrm{P}\left[N_{ht}^{Pj} = n_{ht}^{Pj},\ \ t = 1, 2, \ldots, T,\ \ j \in \{1, 2\}\right] \\
&= \prod_{h \in \mathcal{H}_{12}} \int_0^\infty \int_0^\infty \prod_{t=1}^{T} \left(\exp(-\lambda_{ht}^{P1}\theta_1) \frac{(\lambda_{ht}^{P1}\theta_1)^{n_{ht}^{P1}}}{n_{ht}^{P1}!} \exp(-\lambda_{ht}^{P2}\theta_2) \frac{(\lambda_{ht}^{P2}\theta_2)^{n_{ht}^{P2}}}{n_{ht}^{P2}!} \right) f_{\Theta}(\theta_1, \theta_2) d\theta_1 d\theta_2 \\
\mathcal{L}_2 &= \prod_{h \in \mathcal{H}_{1\backslash 2}} \mathrm{P}\left[N_{ht}^{P1} = n_{ht}^{P1},\ \ t = 1, 2, \ldots, T\right] \\
&= \prod_{h \in \mathcal{H}_{1\backslash 2}} \int_0^\infty \prod_{t=1}^{T} \exp(-\lambda_{ht}^{P1}\theta_1) \frac{(\lambda_{ht}^{P1}\theta_1)^{n_{ht}^{P1}}}{n_{ht}^{P1}!} f_{\Theta^{P1}}(\theta_1) d\theta_1 \\
\mathcal{L}_3 &= \prod_{h \in \mathcal{H}_{2\backslash 1}} \mathrm{P}\left[N_{ht}^{P2} = n_{ht}^{P2},\ \ t = 1, 2, \ldots, T\right] \\
&= \prod_{h \in \mathcal{H}_{2\backslash 1}} \int_0^\infty \prod_{t=1}^{T} \exp(-\lambda_{ht}^{P2}\theta_2) \frac{(\lambda_{ht}^{P2}\theta_2)^{n_{ht}^{P2}}}{n_{ht}^{P2}!} f_{\Theta^{P2}}(\theta_2) d\theta_2.
\end{aligned}$$

The integrals are computed numerically, using the Gauss-Hermite quadrature. A sensitivity analysis to the number of nodes has been conducted.

	$\widehat{\mathrm{V}}[\Theta^{P1}]$		$\widehat{\mathrm{V}}[\Theta^{P2}]$		$\widehat{\rho}_{\mathbf{\Theta}}^{P:P}$	
Moment estimates, aggregated data	0.6025506		0.5260470		0.4901516	
Moment estimates, yearly data	0.6763232		0.6118727		0.3704505	
Bivariate LogNormal Random effect	**0.7221970**		**0.6699701**		**0.4114065**	
95% Confidence interval	0.7175831	0.7268109	0.6454272	0.6945129	0.3924375	0.4303756
	$\widehat{\mathrm{V}}[\Theta^{P1}]$		$\widehat{\mathrm{V}}[\Theta^{P2}]$		R	
Bivariate Gamma Random Effect	0.6630411		0.5884350		0.4349950	
95% Confidence interval	0.6610950	0.6649871	0.5855606	0.5913094	0.4135103	0.4564797

Table: Summary of the estimates along with 95% confidence intervals

The confidence intervals have been computed using the multivariate Delta method, the gradient being estimated numerically.

Vuong Test to compare both models: Test statistic 12.77682, p-value $< 1.11 \times 10^{-37}$

$\Longrightarrow$ Poisson-LogNormal model outperforms the Poisson-Gamma model.

Let us include the Young Drivers in the model

- Let $m_{h,3}$ (resp. $m_{h,4}$) be the number of policyholders from P3 (resp. P4) in household h, i.e. the **number of sons (resp. daughters)** having their own vehicle insured by the company, so that they appear in the databasis.
- Let us now supplement the model for the numbers of claims filed by the parents with additional assumptions to include the claims filed by their children.
  1. For $k \in \{3,4\}$ and $j \in \{1,\ldots,m_{h,k}\}$, given $\Theta_h^{Pk:j} = \theta$, the random variables $N_{h1}^{Pk:j}, N_{h2}^{Pk:j}, \ldots, N_{hT}^{Pk:j}$ are independent, Poisson distributed with respective means $\lambda_{h1}^{Pk:j}\theta, \lambda_{h2}^{Pk:j}\theta, \ldots, \lambda_{hT}^{Pk:j}\theta$.
  2. Given $\Theta_h^{Pk:j}$, $k \in \{3,4\}$ and $j \in \{1,\ldots,m_{h,k}\}$, the sequences $(N_{h1}^{Pk:j}, N_{h2}^{Pk:j}, \ldots, N_{hT}^{Pk:j})$ are independent for different values of k and j.
  3. The random effects $\Theta_h^{Pk:j}$ are LogNormally distributed with unit mean, and independent for different values of h.
  4. Given parents' and children's random effects, the corresponding sequences of yearly numbers of claims are independent.

The **dimension** of the random effects distribution **is equal to the size of the household**. More specifically, we denote the variances of the log of the random effects specific to each sub-population by

$$\mathrm{V}[\log \Theta_h^{Pk}] = (\sigma_{\log \boldsymbol{\Theta}}^{Pk})^2 \text{ for } k \in \{1, 2\}$$

and

$$\mathrm{V}[\log \Theta_h^{Pk:j}] = (\sigma_{\log \boldsymbol{\Theta}}^{Pk})^2 \text{ for } k \in \{3, 4\} \text{ and } j \in \{1, \ldots, m_{h,k}\}.$$

The **correlation matrix** between the log of the random effects is assumed to be of the form

$$\boldsymbol{R} = \left(\begin{array}{cc|cccc}
1 & \rho_{\log \boldsymbol{\Theta}}^{P-P} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \ldots & \rho_{\log \boldsymbol{\Theta}}^{P-CH} \\
\rho_{\log \boldsymbol{\Theta}}^{P-P} & 1 & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \ldots & \rho_{\log \boldsymbol{\Theta}}^{P-CH} \\
\hline
\rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & 1 & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} & \ldots & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} \\
\rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} & 1 & \ldots & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} \\
\ldots & \ldots & \ldots & \ldots & & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} \\
\rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{P-CH} & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} & & \rho_{\log \boldsymbol{\Theta}}^{CH-CH} & 1
\end{array} \right)$$

Modelling TPL on a household level

A posteriori claim frequencies: Including Children in the household

- Estimation by **maximum likelihood** in three steps:
  1. Estimate the "parents' bloc" parameters $\sigma^{P1}_{\log \mathbf{\Theta}}, \sigma^{P2}_{\log \mathbf{\Theta}}, \rho^{P-P}_{\log \mathbf{\Theta}}$;
  2. Estimate the "children's bloc" parameters $\sigma^{P3}_{\log \mathbf{\Theta}}, \sigma^{P4}_{\log \mathbf{\Theta}}, \rho^{CH-CH}_{\log \mathbf{\Theta}}$;
  3. Estimate the final parameter $\rho^{P-CH}_{\log \mathbf{\Theta}}$.
- Finally, the optimization was run again with all 7 parameters, with as initial values the estimates from steps 1-3.

	Estimated Variance	95% Confidence interval	
Θ^{P1}	0.7208686	0.7181652	0.7235721
Θ^{P2}	0.6114775	0.6078304	0.6151247
Θ^{P3}	0.5165708	0.4817677	0.5513740
Θ^{P4}	0.1635426	0.1110731	0.2160122
Pair	Estimated correlation	95% Confidence interval	
$(\Theta^{P1}, \Theta^{P2})$	0.4305379	0.4096889	0.4513869
$(\Theta^{P1}, \Theta^{P3})$	0.2091561	0.1752525	0.2430597
$(\Theta^{P1}, \Theta^{P4})$	0.2188169	0.1834439	0.2541899
$(\Theta^{P2}, \Theta^{P3})$	0.2121019	0.1774682	0.2467356
$(\Theta^{P2}, \Theta^{P4})$	0.2222396	0.1858791	0.2586001
$(\Theta^{P3}, \Theta^{P4})$	0		

Table: Estimated variance-covariance parameters of the random effects in the final model.

Outline

Julien Trufin (joint work with Michel Denuit and Florian Pechon) Université Libre de Bruxelles

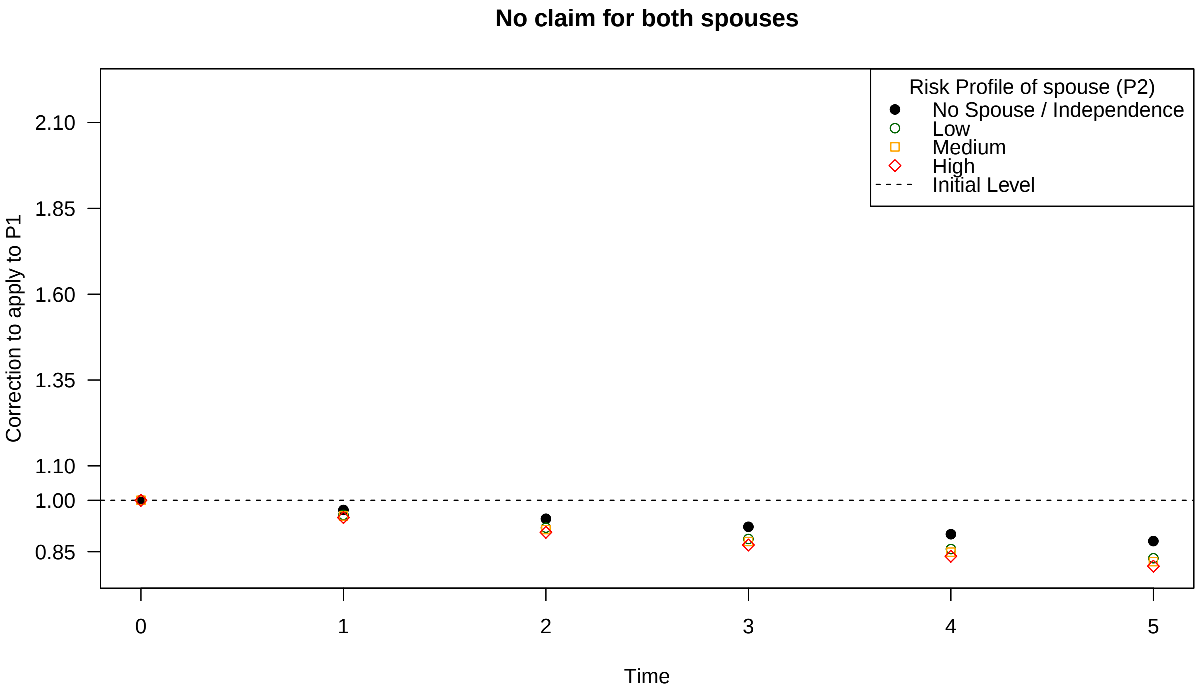
Application 1 : A posteriori corrections

- We can compute the **correction** to apply to each member of the household using all the **information about every policyholder of the household**.
- Using the formula

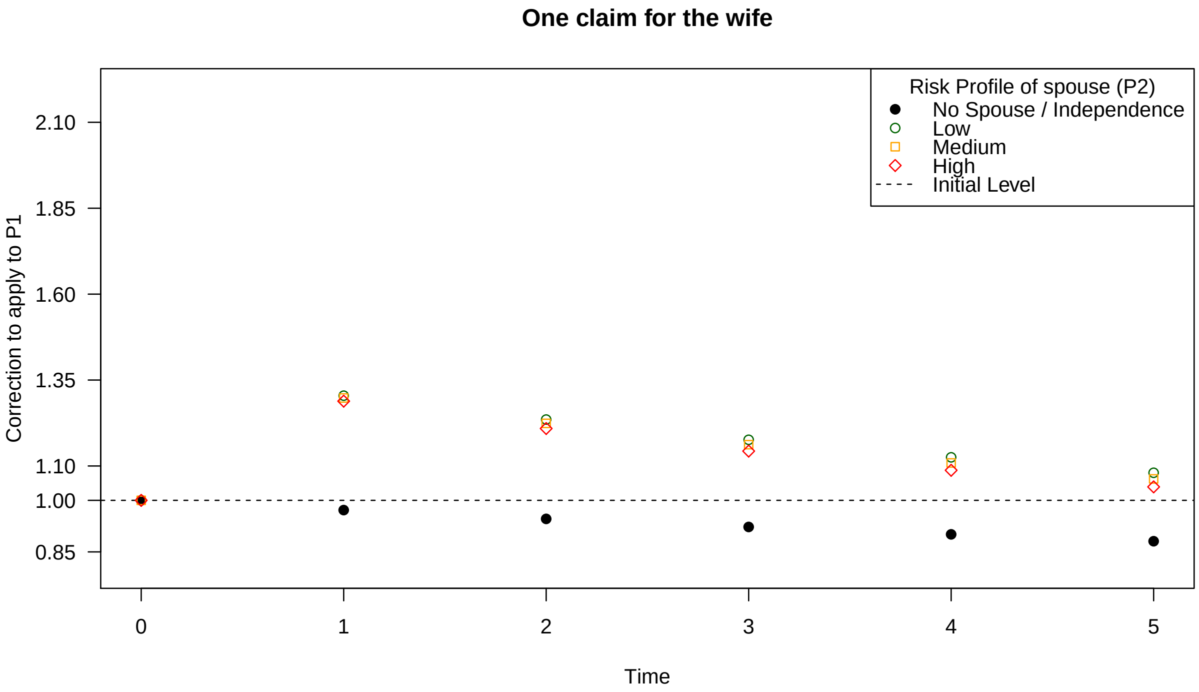
$$E\left[\Theta^{P1}|N_\bullet^{P1}=n_\bullet^{P1},N_\bullet^{P2}=n_\bullet^{P2}\right] = \frac{\int_0^\infty\int_0^\infty \theta^{P1}\exp(-\lambda_\bullet^{P1}\theta^{P1}-\lambda_\bullet^{P2}\theta^{P2})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}\frac{(\lambda_\bullet^{P2}\theta^{P2})^{n_\bullet^{P2}}}{n_\bullet^{P2}!}f_\Theta(\theta^{P1},\theta^{P2})d\theta^{P1}d\theta^{P2}}{\int_0^\infty\int_0^\infty \exp(-\lambda_\bullet^{P1}\theta^{P1}-\lambda_\bullet^{P2}\theta^{P2})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}\frac{(\lambda_\bullet^{P2}\theta^{P2})^{n_\bullet^{P2}}}{n_\bullet^{P2}!}f_\Theta(\theta^{P1},\theta^{P2})d\theta^{P1}d\theta^{P2}}$$

we can compute the correction to apply to the a priori claim frequency of the husband, using his and his wife's claims history.

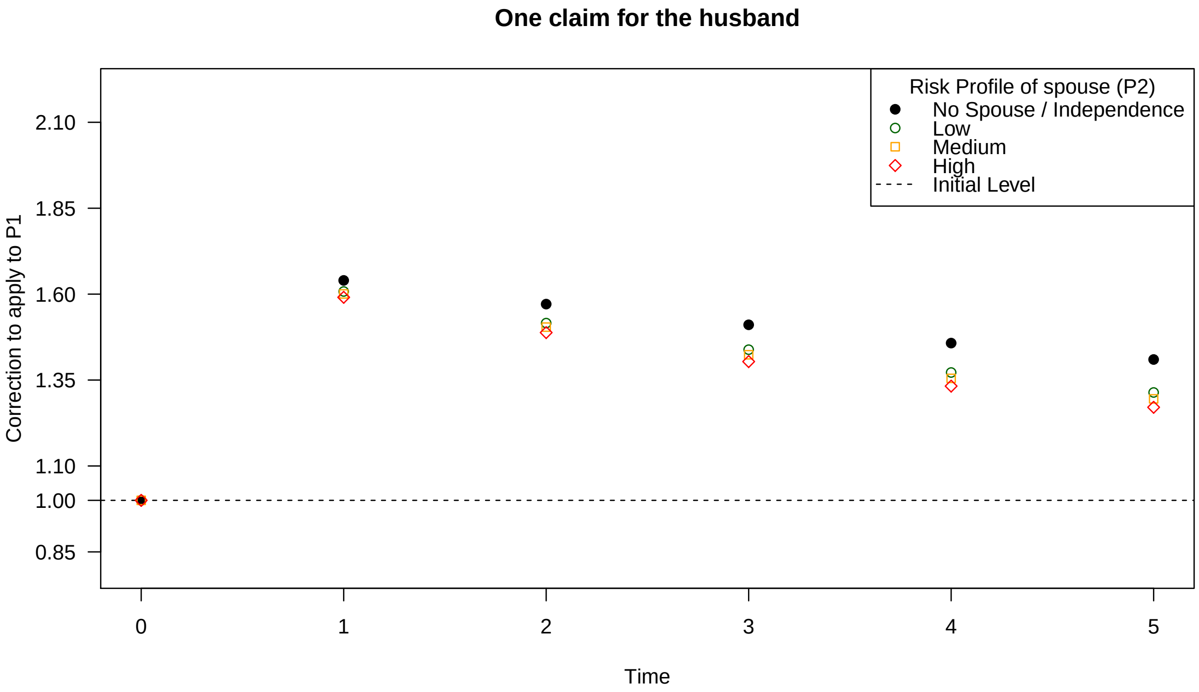
- **Correction to apply to the husband (medium risk profile) when both spouses have had no claim.**



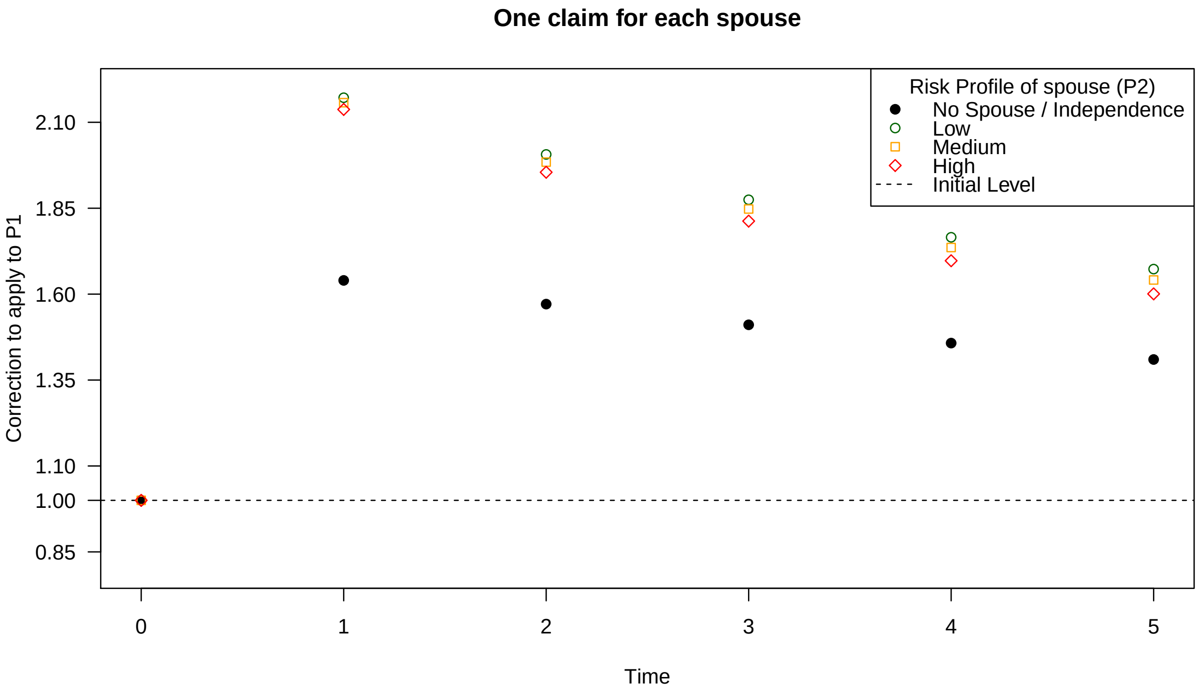
- **Correction to apply to the husband (medium risk profile) when only the wife has had a claim (husband has no claim).**



- Correction to apply to the husband (medium risk profile) when only the husband has had a claim (wife has no claim).



- **Correction to apply to the husband (medium risk profile) when both spouses have had one claim.**



Application 2: Cross-selling

- We can use the multivariate model to perform **cross-selling**.
- For instance, the husband is in the portfolio, but **the wife is not in the portfolio**. **Only information about the husband is available.**

$\Longrightarrow$ We can compute the correction that can be applied to the a priori claim frequency of the wife:

$$E\left[\Theta^{P2}|N_\bullet^{P1}=n_\bullet^{P1}\right]=\frac{\int_0^\infty\int_0^\infty \theta^{P2}\exp(-\lambda_\bullet^{P1}\theta^{P1})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}f_\Theta(\theta^{P1},\theta^{P2})d\theta^{P1}d\theta^{P2}}{\int_0^\infty\exp(\lambda_\bullet^{P1}\theta^{P1})\frac{(-\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}f_\Theta^{P1}(\theta^{P1})d\theta^{P1}}.$$

Recalling that both profiles are clearly associated, we can assume, for instance, that her a priori risk profile level for the past years matches her husband's risk profile.

- We can use the multivariate model to perform **cross-selling**.
- For instance, the husband is in the portfolio, but **the wife is not in the portfolio**. **Only information about the husband is available.**

$\Longrightarrow$ We can compute the correction that can be applied to the a priori claim frequency of the wife.

$n_\bullet^{P1}$	0	1	2	3	4
$E\left[\Theta^{P2} \mid N_\bullet^{P1} = n_\bullet^{P1}\right]$	0.9405	1.1692	1.4319	1.7233	2.0356

Table: Expected value of Θ^{P2} conditional to $N_\bullet^{P1} = n_\bullet^{P1}$, the number of claims of the husband who has been in the portfolio for the past 5 years with a medium level as a priori claim frequency.

- We can use the multivariate model to perform **cross-selling**.
- For instance, the husband is in the portfolio, but **the wife is not in the portfolio**. **Only information about the husband is available.**

$\Longrightarrow$ We can compute the correction that can be applied to the a priori claim frequency of the wife.

$n_\bullet^{P1}$	0	1	2	3	4
$E\left[\Theta^{P2} \mid N_\bullet^{P1} = n_\bullet^{P1}\right]$	0.9405	1.1692	1.4319	1.7233	2.0356

Table: Expected value of Θ^{P2} conditional to $N_\bullet^{P1} = n_\bullet^{P1}$, the number of claims of the husband who has been in the portfolio for the past 5 years with a medium level as a priori claim frequency.

Computations show that $E\left[\Theta^{P2} \mid N_\bullet^{P1} \geq 1\right] = 1.226524$. If we take this as reference level (100%), then the level corresponding to zero claims for the husband is 76.68%.

$\Longrightarrow$ **Wives whose husband had no claim over the past 5 years are around 24-25% less riskier than wives whose husband had at least one claim.**

Application 3 : Underwriting rules for young drivers

- We can use the information about the parents' claims to estimate **correction to apply to the young driver's a priori claim frequency**.
- When **both parents are in the portfolio**, we can use

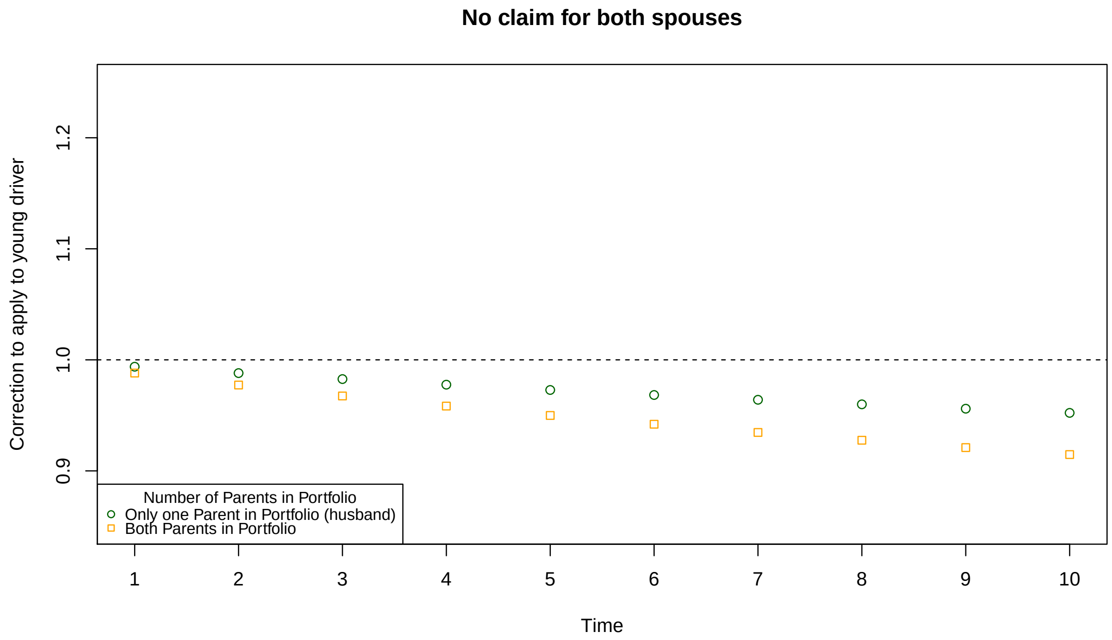
$$E\left[\Theta^{P3}|N_\bullet^{P1}=n_\bullet^{P1},N_\bullet^{P2}=n_\bullet^{P2}\right]$$
$$=\frac{\int_0^\infty\int_0^\infty\int_0^\infty \theta^{P3}\exp(-\lambda_\bullet^{P1}\theta^{P1}-\lambda_\bullet^{P2}\theta^{P2})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}\frac{(\lambda_\bullet^{P2}\theta^{P2})^{n_\bullet^{P2}}}{n_\bullet^{P2}!}f_\Theta(\theta^{P1},\theta^{P2},\theta^{P3})d\theta^{P1}d\theta^{P2}d\theta^{P3}}{\int_0^\infty\int_0^\infty \exp(-\lambda_\bullet^{P1}\theta^{P1}-\lambda_\bullet^{P2}\theta^{P2})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}\frac{(\lambda_\bullet^{P2}\theta^{P2})^{n_\bullet^{P2}}}{n_\bullet^{P2}!}f_\Theta(\theta^{P1},\theta^{P2})d\theta^{P1}d\theta^{P2}}$$

whereas when **only the husband is in the portfolio**, we use

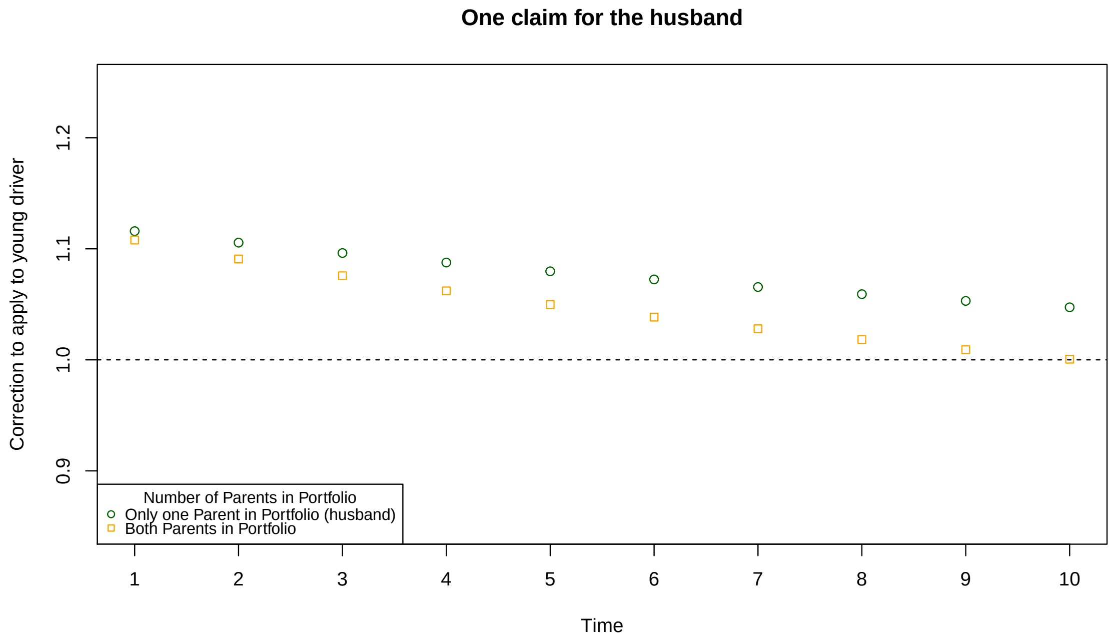
$$E\left[\Theta^{P3}|N_\bullet^{P1}=n_\bullet^{P1}\right]=\frac{\int_0^\infty\int_0^\infty \theta^{P3}\exp(-\lambda_\bullet^{P1}\theta^{P1})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}f_\Theta(\theta^{P1},\theta^{P3})d\theta^{P1}d\theta^{P3}}{\int_0^\infty \exp(-\lambda_\bullet^{P1}\theta^{P1})\frac{(\lambda_\bullet^{P1}\theta^{P1})^{n_\bullet^{P1}}}{n_\bullet^{P1}!}f_\Theta(\theta^{P1})d\theta^{P1}}.$$

- This may be used as an **underwriting rule** for young drivers: Only accept the young drivers whose estimate is below some level. This also shows the usefulness of having both parents in the portfolio.

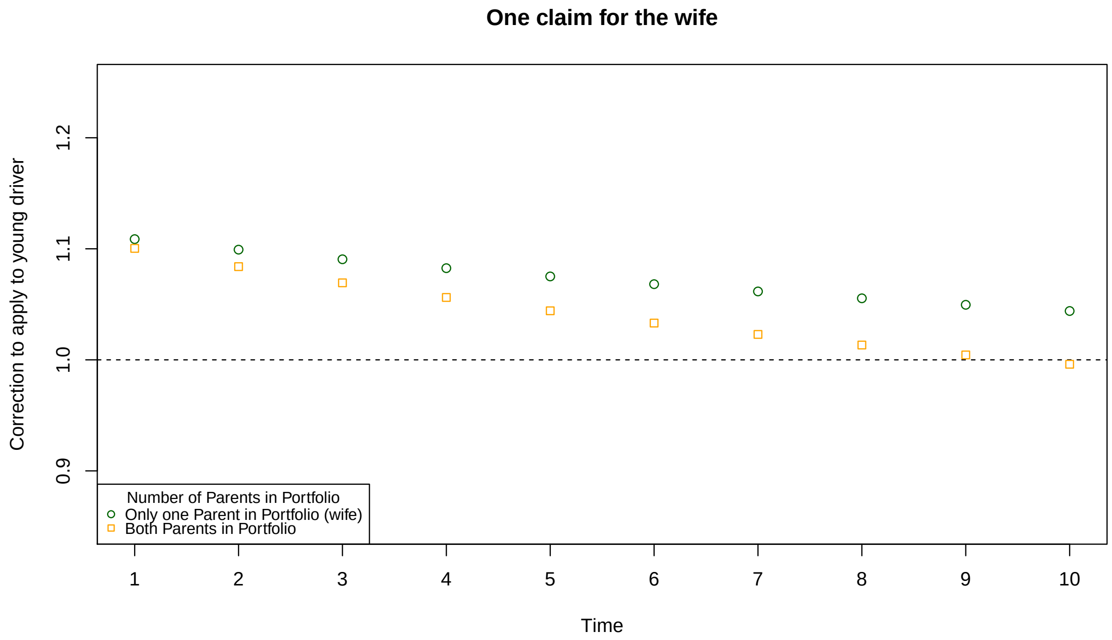
- **Expectation of Θ^{P3} conditionally to the number of claims of the father (green circle) and of both parents (orange square) throughout time (in years).**



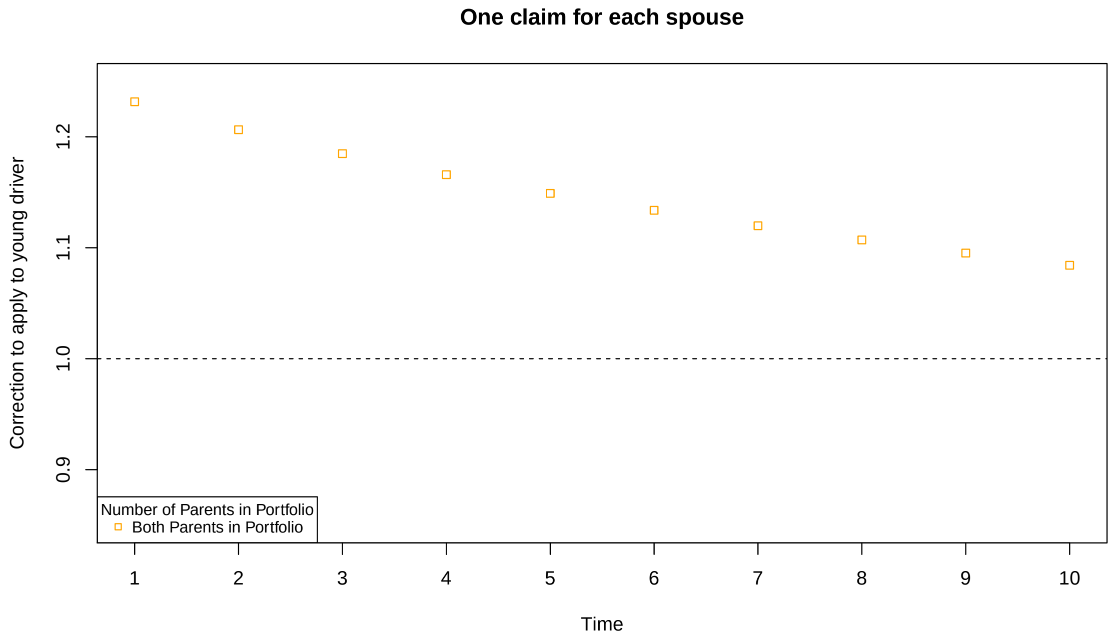
- **Expectation of Θ^{P3} conditionally to the number of claims of the father (green circle) and of both parents (orange square) throughout time (in years).**



- **Expectation of Θ^{P3} conditionally to the number of claims of the wife (green circle) and of both parents (orange square) throughout time (in years).**



- **Expectation of Θ^{P3} conditionally to the number of claims of both parents (orange square) throughout time (in years).**



Reference

- Florian Pechon, Julien Trufin and Michel Denuit (2018), Multivariate modelling of household claim frequencies in motor third-party liability insurance, Accepted in ASTIN Bulletin, to appear.