Multivariate claim frequencies modelling in P&C insurance using multiple guarantees and multiple policyholders from the same household

Florian Pechon, Michel Denuit, Julien Trufin

April 5, 2019

LiDAM, UCLouvain (Belgium)

Hosted by ACTUARIAL SOCIETY OF SOUTH AFRICA

SECTION COLLOQUIUM 2019

Agenda

1. Introduction to ratemaking
2. Proposed Model
3. How can we apply this model ?
4. Conclusions

Introduction

- Insurance allows a policyholder i **to transfer a risk** X_i in exchange of a premium $\pi_i = \mathbb{E}[X_i] +$ a safety margin (and fees & taxes). The **technical premium** is defined as $\mathbb{E}[X_i]$.
- As opposed to other businesses, in insurance, the production cycle is inverted: the insurer does not know the value of the contract he sells at time $t = 0$. Only at the end of the year, the value will be known.
- Often, X_i will in fact be a **compound distribution** (e.g. $X_i = \sum_{j=1}^{N_i} Y_j$, where N_i is a variable counting the number of claims and Y_j is the cost of the claim j).
  $\rightarrow$ The focus in this talk will be on the **claim frequency** $\lambda_i = \mathbb{E}[N_i]$.
- Risk classification generally confined to univariate analysis: Each product is considered in isolation.
- Independence between policyholders is generally assumed.
- Generally, consists of two parts:
  1. **A priori risk classification**: Use the **available covariates** (e.g. Age of the driver, place of residence, split of the premium) to classify the policyholders into risk classes. Policyholder belonging to the same risk class are given the same claim frequency.
  2. **A posteriori risk classification**: As time passes, the claim experience reveals information about some **hidden, latent risk factors**. The a priori claim frequency is adapted thanks to **credibility theory**.

Introduction

The two main guarantees in Motor insurance are **Third-Party Liability (TPL)** insurance and **Material Damage (MD)** insurance.

- TPL is compulsory and covers a third-party's loss caused by the insured car.
- MD is an optional guarantee that covers the cost of repairing or replacing the insured's own vehicle. The policyholder will typically trigger this guarantee when he is liable for the claim, or could not identify the liable person.

Main Idea of this presentation: Show how we can estimate the claim frequencies in both guarantees while **taking into account the dependencies** that may exist:

- between both guarantees;
- between policyholders from the same household.

Model

Model

1. **Dependence between guarantees :** Due to the nature of the guarantees, one single event can sometimes trigger both guarantees at the same time.

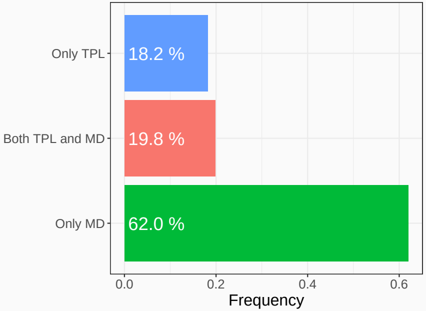


Figure 1: Relative occurrence according to the type of claims.

2. **Dependence between policyholders :** Some latent (unobserved) important risk factors are not observed (e.g. driving in dangerous conditions). These latent risk factors may be **shared across guarantees and/or policyholders** from the same household.

Model: a priori risk classification

Let us introduce the following claim count variables :

- $N_{h(i),t}^{TPL}$: Number of claims of policyholder i from household h that triggered **only TPL** during year t;
- $N_{h(i),t}^{MD}$: Number of claims of policyholder i from household h that triggered **only MD** during year t;
- $N_{h(i),t}^{MD:TPL}$: Number of claims of policyholder i from household h that triggered **both TPL and MD simultaneously** during year t.

The **total the number of claims** for policyholder i from household h during year t that trigger

- TPL is $N_{h(i),t}^{TPL} + N_{h(i),t}^{MD:TPL}$;
- MD is $N_{h(i),t}^{MD} + N_{h(i),t}^{MD:TPL}$.

The corresponding a priori claim frequencies are given by

$$\begin{cases} \lambda_{h(i),t}^{TPL} &= \mathbb{E}\left[N_{h(i),t}^{TPL}\right] \\ \lambda_{h(i),t}^{MD} &= \mathbb{E}\left[N_{h(i),t}^{MD}\right] \\ \lambda_{h(i),t}^{MD:TPL} &= \mathbb{E}\left[N_{h(i),t}^{MD:TPL}\right] \end{cases}$$

The claim frequencies can be estimated with a Poisson regression, for instance, using GAMs (Generalized Additive Models).

$\Longrightarrow$ We account for claims that trigger both guarantees at the same time.

Introduce **random effects** that account for unobserved heterogeneity (*over-dispersion*): Some important risk factors are not observed; they will be represented by these random effects.

$$\begin{cases} \begin{cases} \lambda^{TPL}_{h(1),\bullet}\Theta^{TPL}_{h(1)} & = \exp(\boldsymbol{X}'\beta)\Theta^{TPL}_{h(1)} = \exp(\boldsymbol{X}'\beta + \epsilon^{TPL}_{h(1)}) \\ \lambda^{MD}_{h(1),\bullet}\Theta^{MD}_{h(1)} & = \exp(\boldsymbol{X}'\beta)\Theta^{MD}_{h(1)} = \exp(\boldsymbol{X}'\beta + \epsilon^{MD}_{h(1)}) \\ \lambda^{MD:TPL}_{h(1),\bullet}\Theta^{MD:TPL}_{h(1)} & = \exp(\boldsymbol{X}'\beta)\Theta^{MD:TPL}_{h(1)} = \exp(\boldsymbol{X}'\beta + \epsilon^{MD:TPL}_{h(1)}) \end{cases} \\ \\ \begin{cases} \lambda^{TPL}_{h(2),\bullet}\Theta^{TPL}_{h(2)} & = \exp(\boldsymbol{X}'\beta)\Theta^{TPL}_{h(2)} = \exp(\boldsymbol{X}'\beta + \epsilon^{TPL}_{h(2)}) \\ \lambda^{MD}_{h(2),\bullet}\Theta^{MD}_{h(2)} & = \exp(\boldsymbol{X}'\beta)\Theta^{MD}_{h(2)} = \exp(\boldsymbol{X}'\beta + \epsilon^{MD}_{h(2)}) \\ \lambda^{MD:TPL}_{h(2),\bullet}\Theta^{MD:TPL}_{h(2)} & = \exp(\boldsymbol{X}'\beta)\Theta^{MD:TPL}_{h(2)} = \exp(\boldsymbol{X}'\beta + \epsilon^{MD:TPL}_{h(2)}) \end{cases} \end{cases}$$

Let $\boldsymbol{\Theta_h} = (\Theta^{TPL}_{h(1)}, \Theta^{MD}_{h(1)}, \Theta^{MD:TPL}_{h(1)}, \Theta^{TPL}_{h(2)}, \Theta^{MD}_{h(2)}, \Theta^{MD:TPL}_{h(2)})$. We will assume that $\boldsymbol{\Theta_h}$ has a **multivariate LogNormal distribution**.

Let $\boldsymbol{\Theta_h} = (\Theta_{h(1)}^{TPL}, \Theta_{h(1)}^{MD}, \Theta_{h(1)}^{MD:TPL}, \Theta_{h(2)}^{TPL}, \Theta_{h(2)}^{MD}, \Theta_{h(2)}^{MD:TPL})$.

- We need $\mathbb{E}[\boldsymbol{\Theta_h}] = \mathbf{1}$.
- We impose that
  - $\mathbb{V}\left[\log \Theta_{h(i)}^{g}\right] = \sigma_g^2 \; \forall g \in \mathcal{G} := \{TPL, MD, MD : TPL\}$ and $i = 1, 2$.
  -

$$\mathbb{C}orr[\log \boldsymbol{\Theta_h}] = \left(\begin{array}{ccc|ccc} 1 & \rho^{TPL,MD} & \rho^{TPL,MD:TPL} & \rho_{12} & \rho_{12} & \rho_{12} \\ \rho^{TPL,MD} & 1 & \rho^{MD,MD:TPL} & \rho_{12} & \rho_{12} & \rho_{12} \\ \rho^{TPL,MD:TPL} & \rho^{MD,MD:TPL} & 1 & \rho_{12} & \rho_{12} & \rho_{12} \\ \hline \rho_{12} & \rho_{12} & \rho_{12} & 1 & \rho^{TPL,MD} & \rho^{TPL,MD:TPL} \\ \rho_{12} & \rho_{12} & \rho_{12} & \rho^{TPL,MD} & 1 & \rho^{MD,MD:TPL} \\ \rho_{12} & \rho_{12} & \rho_{12} & \rho^{TPL,MD:TPL} & \rho^{MD,MD:TPL} & 1 \end{array}\right).$$

We rely on **maximum likelihood** to estimate the variance-covariance matrix.

	Estimate
$\widehat{\mathbb{V}}(\log \Theta^{TPL})$	0.55843
$\widehat{\mathbb{V}}(\log \Theta^{MD})$	0.36473
$\widehat{\mathbb{V}}(\log \Theta^{MD:TPL})$	0.31750
$\widehat{\mathbb{C}orr}(\log \Theta^{TPL}, \log \Theta^{MD})$	0.52414
$\widehat{\mathbb{C}orr}(\log \Theta^{TPL}, \log \Theta^{MD:TPL})$	0.69405
$\widehat{\mathbb{C}orr}(\log \Theta^{MD}, \log \Theta^{MD:TPL})$	0.51272
$\widehat{\rho}$	0.44117

Table 1: Maximum likelihood estimates of the variances and correlations of the underlying Normal random variables, i.e. $\log \boldsymbol{\Theta}$, where the log is taken on each component of the vector.

Applications

Applications

Some applications are displayed hereafter.

- Correct the a priori claim frequency.
- Identify profitable (new) policyholders using the past claims of the rest of the household.

Correct the a priori claim frequency

Let us show the impact of claims on the a priori claim frequency.

We assume that we have observed a household with **a single policyholder**, with a median a priori risk profile, over the past T years.

The conditional expectations of the random effects can be computed using the **Bayes formula** and numerical integration.

We can compute the ratios

$$\mathbb{E}\left[\frac{\lambda_{h(1),T+1}^{TPL}\Theta_{h(1)}^{TPL}+\lambda_{h(1),T+1}^{MD:TPL}\Theta_{h(1)}^{MD:TPL}}{\lambda_{h(1),T+1}^{TPL}+\lambda_{h(1),T+1}^{MD:TPL}}|N_{h(1),\bullet}^{\tilde{g}}=n_{h(1),\bullet}^{\tilde{g}},\forall\tilde{g}\in\mathcal{G}\right]$$

and

$$\mathbb{E}\left[\frac{\lambda_{h(1),T+1}^{MD}\Theta_{h(1)}^{MD}+\lambda_{h(1),T+1}^{MD:TPL}\Theta_{h(1)}^{MD:TPL}}{\lambda_{h(1),T+1}^{MD}+\lambda_{h(1),T+1}^{MD:TPL}}|N_{h(1),\bullet}^{\tilde{g}}=n_{h(1),\bullet}^{\tilde{g}},\forall\tilde{g}\in\mathcal{G}\right]$$

Correction factors when no claim is reported

We can compute the correction to apply to policyholder $h(1)$. We assume

- **No claim** has been reported in any of the two guarantees.
- $h(1)$ is the only policyholder from the household.

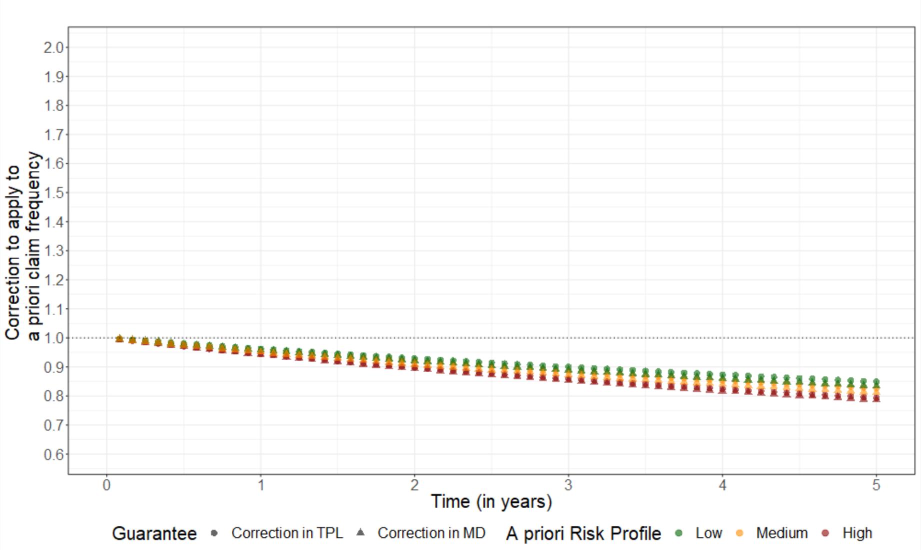


Figure 2: Correction to apply to TPL and MD for a household with a single policyholder. No claim in any guarantee.

Correction factors after a claim in TPL

We can compute the correction to apply to policyholder $h(1)$. We assume

- **One claim** triggering **TPL only** occurred at some time t.
- $h(1)$ is the only policyholder from the household.

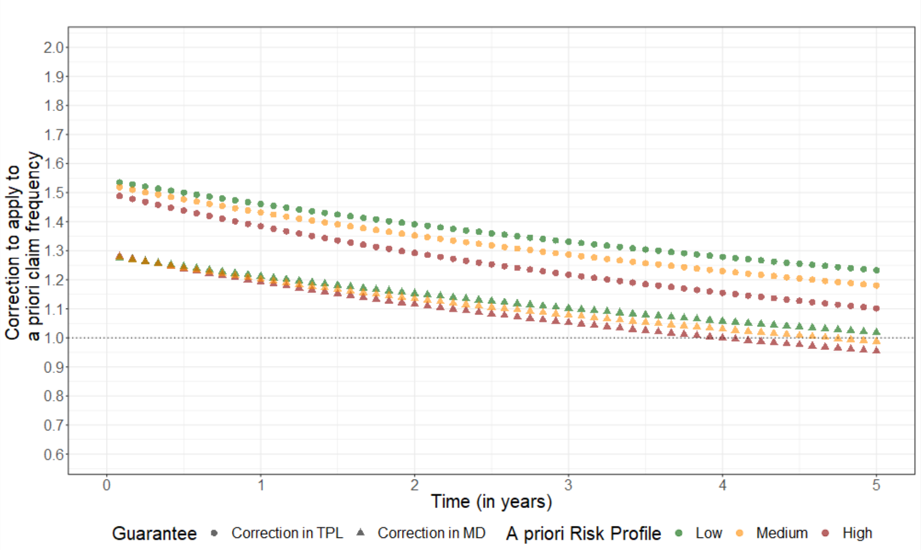


Figure 3: Correction to apply to TPL and MD for a household with a single policyholder. No claim in any other guarantee.

Correction factors after a claim in MD

We can compute the correction to apply to policyholder $h(1)$. We assume

- **One claim** triggering **MD only** occurred at some time t.
- $h(1)$ is the only policyholder from the household.

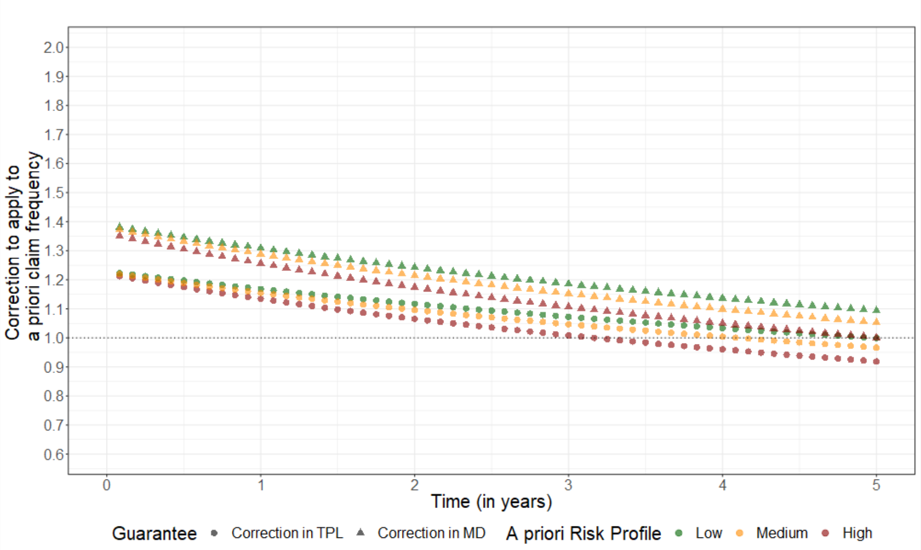


Figure 4: Correction to apply to TPL and MD for a household with a single policyholder. No claim in any other guarantee.

One single event or two claims ?

We assume that the policyholder has a **median** a priori risk profile.

We compare two cases:

$$\begin{cases} (N^{TPL}_{h(1),\bullet}, N^{MD}_{h(1),\bullet}, N^{MD:TPL}_{h(1),\bullet}) = (0,0,1) \\ (N^{TPL}_{h(1),\bullet}, N^{MD}_{h(1),\bullet}, N^{MD:TPL}_{h(1),\bullet}) = (1,1,0) \end{cases}$$

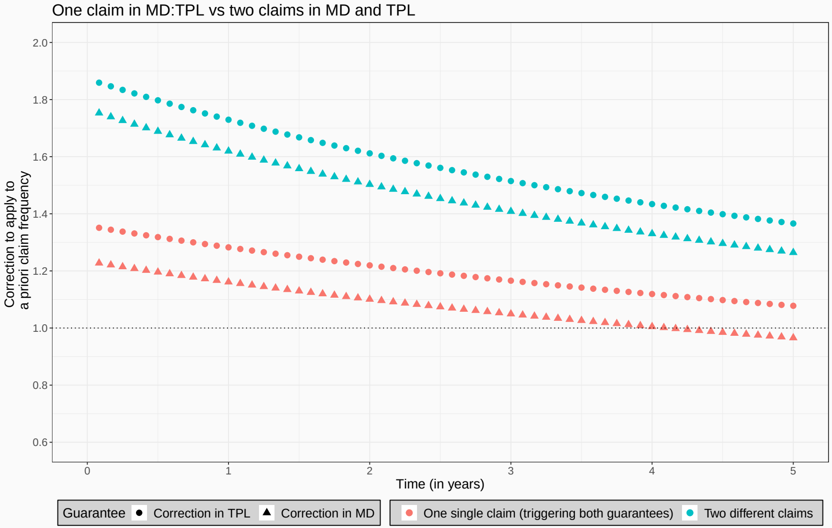


Figure 5: Correction to apply to TPL and MD for a household with a single policyholder. Comparison when a claim triggered both guarantees with two claims (one in each guarantee).

Identify cross-selling opportunities

Identify **new (profitable) policyholders** thanks to informations related to the rest of the household.

We can compute the corrections that we could apply to these new policyholders.

Identify cross-selling opportunities

Identify **new (profitable) policyholders** thanks to informations related to the rest of the household.

We can compute the corrections that we could apply to these new policyholders.

$$\mathbb{E}\left[\lambda^{TPL}_{h(\star),T+1}\Theta^{TPL}_{h(\star)} + \lambda^{MD:TPL}_{h(\star),T+1}\Theta^{MD:TPL}_{h(\star)} | N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]$$

$$= \lambda^{TPL}_{h(\star),T+1} \frac{\int_0^\infty \cdots \int_0^\infty \theta^{TPL}_{h(\star)} P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G} | \boldsymbol{\Theta_h} = \boldsymbol{\theta}\right] f_{\boldsymbol{\Theta_h}}(\boldsymbol{\theta})}{P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]} + \lambda^{MD:TPL}_{h(\star),T+1} \frac{\int_0^\infty \cdots \int_0^\infty \theta^{MD:TPL}_{h(\star)} P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G} | \boldsymbol{\Theta_h} = \boldsymbol{\theta}\right] f_{\boldsymbol{\Theta_h}}(\boldsymbol{\theta})}{P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]}$$

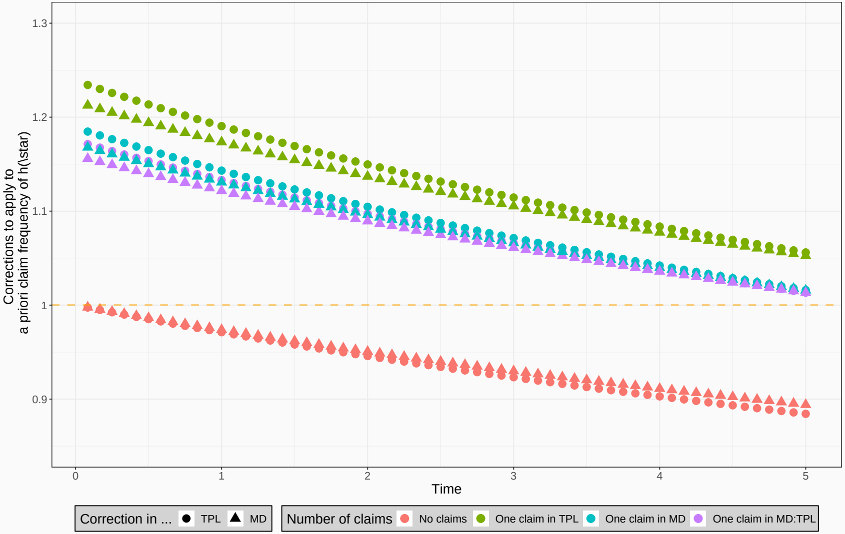
and

$$\mathbb{E}\left[\lambda^{MD}_{h(\star),T+1}\Theta^{MD}_{h(\star)} + \lambda^{MD:TPL}_{h(\star),T+1}\Theta^{MD:TPL}_{h(\star)} | N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]$$

$$= \lambda^{MD}_{h(\star),T+1} \frac{\int_0^\infty \cdots \int_0^\infty \theta^{MD}_{h(\star)} P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G} | \boldsymbol{\Theta_h} = \boldsymbol{\theta}\right] f_{\boldsymbol{\Theta_h}}(\boldsymbol{\theta})}{P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]} + \lambda^{MD:TPL}_{h(\star),T+1} \frac{\int_0^\infty \cdots \int_0^\infty \theta^{MD:TPL}_{h(\star)} P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G} | \boldsymbol{\Theta_h} = \boldsymbol{\theta}\right] f_{\boldsymbol{\Theta_h}}(\boldsymbol{\theta})}{P\left[N^{g}_{h(1),\bullet} = n^{g}_{h(1),\bullet} \forall g \in \mathcal{G}\right]}.$$

Identify cross-selling opportunities

- We can discuss the number of claims reported by $h(1)$.
- We assume that we don't have any information related to the past claims experience of $h(\star)$.
- We assume a **median** a priori risk profile.



Conclusions

Conclusions

- We showed how to account for the dependencies between the guarantees and the policyholders in motor insurance.
- Dependency between guarantees arises from two different aspects:
  - One claim can trigger both guarantees at the same time;
  - Latent (unobserved) risk factors affecting the claim frequency appear to be correlated.
- Dependency between policyholders comes from correlated latent (unobserved) risk factors.

→ Paper is currently in revision.

- Other applications of similar model in Motor TPL: Consider a Family in a Household:
  - Integrate two different kind of policyholders: Adults (or Parents) and Young Drivers
  - Consider different correlations depending on the type of policyholders (Adult or Young Drivers).

  → Paper is published in the *ASTIN Bulletin*.

Florian Pechon, Michel Denuit, and Julien Trufin.
Multivariate modelling of multiple guarantees in motor insurance of a household.
In Revision, 2019.

Florian Pechon, Julien Trufin, and Michel Denuit.
Multivariate modelling of household claim frequencies in motor third-party liability insurance.
ASTIN Bulletin, 48(3):969993, 2018.

Thank you!