

Article

Macro *vs.* Micro Methods in Non-Life Claims Reserving (an Econometric Perspective)

Arthur Charpentier ^{1,2} and Mathieu Pigeon ^{1,*}

¹ Quantact/Département de mathématiques, Université du Québec à Montréal, Montreal, QC H2X 3Y7, Canada; charpentier.arthur@uqam.ca

² Département de Sciences Économiques, Université de Rennes 1, Rennes 35000, France

* Correspondence: pigeon.mathieu.2@uqam.ca; Tel.: +1-514-987-3000 (ext. 2807)

Academic Editor: Montserrat Guillén

Received: 29 February 2016; Accepted: 2 May 2016; Published: 14 May 2016

Abstract: Traditionally, actuaries have used run-off triangles to estimate reserve (“macro” models, on aggregated data). However, it is possible to model payments related to individual claims. If those models provide similar estimations, we investigate uncertainty related to reserves with “macro” and “micro” models. We study theoretical properties of econometric models (Gaussian, Poisson and quasi-Poisson) on individual data, and clustered data. Finally, applications in claims reserving are considered.

Keywords: loss reserving; clustering; generalized linear mixed models

1. Introduction

1.1. Macro and Micro Methods

For more than a century, actuaries have been using run-off triangles to project future payments, in non-life insurance. In the 1930s, [1] formalized this technique that originated from the popular *chain ladder* algorithm. In the 1990s, [2] proved that the chain ladder estimate can be motivated by a simple stochastic model, and later on [3] provided a comprehensive overview on stochastic models that can be connected with the *chain ladder* method, including regression models that could be seen as extension of the so-called “factor” methods used in the 1970s.

The terminology of [4] and [5] were used in macro-level models for reserving. In the 1970s, [6] suggested using some marked point process of claims to project future payments, and quantify the reserves. More recently, [7–10] or [11] (among many others) investigated further some probabilistic micro-level models. These models handle claims related data on an individual basis, rather than aggregating by underwriting year and development period. As mentioned in [12], these methods have not (yet) found great popularity in practice, since they are more difficult to apply. Nevertheless, several papers address that issue, with several stochastic processes to model the dynamics of payments, such as [13]—extended in [14,15] or [16], and more recently [17] and [18].

All macro-level models are based on aggregate data found in a run-off triangle, which is their strength, but also probably their weakness. Indeed, with regulation rules, such as *Solvency II* or *IFRS 4—Phase 2* norms, the goal is no longer to get (only) a best-estimate, but it becomes necessary to have the full conditional distribution of future cash-flows.

Macro-level approaches, such as the popular *chain ladder* model, have undeniable properties. Those models are part of the actuarial folklore and extensions have been derived to have more general and more realistic models. They are easy to understand, and can be mentioned in financial communication, without disclosing too much information. From a computational perspective,

those models can also be implemented in a single spreadsheet and more advanced libraries, such as the ChainLadder package in R can also be used.

Nevertheless, those macro-level models have also major drawbacks, as discussed for instance in [16] or [17], and references therein. More specifically, we wish to highlight two important points:

- Those models neglect a lot of information that is available on a micro-level (per individual claim). Some additional covariates can be used, as well as exposure, *etc.* In most applications, not only is that information available, but usually, it has a valuable predictive power. To use that additional information, one cannot simply modify macro-level models, and it is necessary to change the general framework of the model. It becomes possible to emphasize large losses and to distinguish them from regular claims, to get more detailed information about future payments, *etc.*
- As discussed in this paper, macro-level models on aggregated data can be seen as models on clusters and not on individual observations, as we will do with micro-level models. In the context of macro-level models for loss reserving, [3] mention that prediction errors can be large, because of the small number of observations used in run-off triangles and the fact that clusters are usually not homogeneous. Quantifying uncertainty in claim reserving methods is not only important in actuarial practice and to assess accuracy of predictive models, it is also a regulatory issue. Finally, a small sample size can cause a lack of robustness and a risk of over-parametrization for macro-level models.

On the other hand, micro-level models have many pros. They can incorporate micro-level covariates, and model micro-structure of claims development. Thus, they can take into account heterogeneity, structural changes, *etc.* (see [19] for a discussion). Several studies, with empirical comparisons between macro- and micro-level models (see, e.g., [17]), illustrate that in many scenarios, reserves obtained with micro-level models have higher precision than macro-level ones. Further, since those models are fitted on much more data than aggregated ones, variability of predictions is usually smaller, which increases model robustness and reduces the risk of over-fitting. More specifically, [15] and [16] obtained, on real data analysis, lower variance on the total amount of reserves with “micro” models than with “macro” ones. A natural question is about the generality of such result. Should “micro” model generate less variability than standard “macro” ones? That is the question that initiated this paper.

1.2. Agenda

In Section 2, we detail intuitive results we expect when aggregating data by clusters, moving from micro-level models to macro-level ones. More precisely, we explain why with a linear model and a Poisson regression, macro- and micro-level models are equivalent. We also discuss the case of the Poisson regression model with random intercept. In Section 3, we study “micro” and “macro” models in the context of claims reserving, on real data, as well as simulated ones. More specifically, we present a methodology that allows to generate—by random generation—some micro-datasets obtained from macro-datasets. We investigate also the impact of adding micro-type covariates, allowing for various types of correlations. Finally, we present concluding remarks in Section 4.

2. Clustering in Generalized Linear Mixed Models

In the economic literature, several papers discuss the use of “micro” *vs.* “macro” data, for instance in the context of unemployment duration in [20] or in the context of inflation in [21]. In [20], it is mentioned that both models are interesting, since “micro” data can be used to capture heterogeneity while “macro” data can capture cycle and more structural patterns. In [21], it is demonstrated that both heterogeneity and aggregation might explain the persistence of inflation at the macroeconomic level.

In order to clarify notation, and make sure that objects are well defined, we use small letters for sample values, e.g., y_i , and capital letters for underlying random variables, e.g., Y_i in the sense that y_i is a realisation of random variable Y_i . Hence, in the case of the linear model (see Subsection 2.1), we usually assume that $Y_i \sim \mathcal{N}(x_i^T \mathbf{b}, \sigma^2)$, and then $\hat{\mathbf{b}}$ is the vector of estimated parameters, in the sense

that $\hat{\mathbf{b}} = (\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{y}$ while $\hat{\mathbf{B}} = (\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{Y}$ (here covariates \mathbf{x} are given, and non stochastic). Since $\hat{\mathbf{B}}$ is seen as a random variable, we can write $\mathbb{E}[\hat{\mathbf{B}}] = \mathbf{b}$.

With a Poisson regression, $Y_i \sim \mathcal{P}(\lambda_i)$ with $\lambda_i = \exp[\mathbf{x}_i^\top \mathbf{b}]$. In that case, $\text{Var}[Y_i] = \mathbb{E}[Y_i] = \lambda_i$. The estimate parameter $\hat{\mathbf{b}}$ is a function of the observations, (\mathbf{x}_i, y_i) 's, while $\hat{\mathbf{B}}$ is a function of the observations \mathbf{x}_i and of the random variable Y_i . In the context of the Poisson regression, recall that $\mathbb{E}[\hat{\mathbf{B}}] \rightarrow \mathbf{b}$ as n , the number of observations, goes to infinity. With a quasi-Poisson regression, Y_i does not have, per se, a proper distribution. Nevertheless, its moments are well defined, in the sense that $\text{Var}[Y_i] = \varphi \mathbb{E}[Y_i] = \varphi \lambda_i$, where φ denotes the dispersion parameter (see Subsection 2.2). For convenience, we will denote $Y_i \sim q\mathcal{P}(\lambda_i)$, with an abuse of notation.

In this section, we will derive some theoretical results regarding aggregation in econometric models.

2.1. The Multiple Linear Regression Model

For a $(k + 1) \times 1$ vector of parameters \mathbf{a} , we consider a (multiple) linear regression model,

$$y_{i,g} = \mathbf{x}_g^\top \mathbf{a} + \varepsilon_{i,g} \tag{1}$$

$$\mathbf{a} = [a_1 \quad \dots \quad a_{k+1}]^\top \quad \mathbf{x}_g = [x_{g,1} \quad \dots \quad x_{g,k+1}]^\top$$

where observations belong to a cluster g and are indexed by i within a cluster g , $i = 1, \dots, n_g$, $g = 1, \dots, m$. Assume further assumptions of the classical linear regression model [22], i.e.,

- (LRM1) no multicollinearity in the data matrix;
- (LRM2) exogeneity of the independent variables $\mathbb{E}[\varepsilon_{i,g} | \mathbf{x}_g] = 0$, $i = 1, \dots, n_g$, $g = 1, \dots, m$; and
- (LRM3) homoscedasticity and nonautocorrelation of error terms with $\text{Var}[\varepsilon_{i,g}] = \sigma^2$.

Stacking observations within a cluster yield the following model

$$\bar{y}_g = \mathbf{x}_g^\top \mathbf{b} + e_g \tag{2}$$

where

$$\bar{y}_g = \frac{1}{n_g} \sum_i y_{i,g} \text{ and } \mathbf{b} = [b_1 \quad \dots \quad b_{k+1}]^\top$$

with similar assumptions except for $\text{Var}[e_g] = \sigma^2/n_g$. Those two models are equivalent, in the sense that the following proposition holds.

Proposition 1. *Model (1) on a micro level and Model (2) on a macro level are equivalent, in the sense that*

- (i) $\hat{\mathbf{a}}_{OLS} = \hat{\mathbf{b}}_{OLS}$ when weights n_g are used in Model (2); and
- (ii) $\sum_{i,g} \hat{y}_{i,g} = \sum_g \hat{y}_g$ where $y_g = n_g \bar{y}_g$.

Proof. (i) The ordinary least-squares estimator for \mathbf{a} - from Model (1)—is defined as

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\text{argmin}} \left\{ \sum_{i,g} (y_{i,g} - \mathbf{x}_g^\top \mathbf{a})^2 \right\} \tag{3}$$

which can also be written

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\text{argmin}} \left\{ \sum_{i,g} (y_{i,g} - \bar{y}_g + \bar{y}_g - \mathbf{x}_g^\top \mathbf{a})^2 \right\}. \tag{4}$$

Now, observe that

$$\sum_{i,g} \left(y_{i,g} - \bar{y}_g + \bar{y}_g - \mathbf{x}_g^T \mathbf{a} \right)^2 = \sum_{i,g} (y_{i,g} - \bar{y}_g)^2 + (\bar{y}_g - \mathbf{x}_g^T \mathbf{a})^2 + 2(y_{i,g} - \bar{y}_g)(\bar{y}_g - \mathbf{x}_g^T \mathbf{a})$$

where the first term is independent of \mathbf{a} (and can be removed from the optimization program), and the term with cross-elements sums to 0. Hence,

$$\hat{\mathbf{a}} = \operatorname{argmin}_a \left\{ \sum_{i,g} (\bar{y}_g - \mathbf{x}_g^T \mathbf{a})^2 \right\} = \operatorname{argmin}_a \left\{ \sum_g n_g (\bar{y}_g - \mathbf{x}_g^T \mathbf{a})^2 \right\} = \hat{\mathbf{b}} \tag{5}$$

where $\hat{\mathbf{b}}$ is the least square estimator of \mathbf{b} from Model (2), when weights n_g are considered.

(ii) If we consider the sum of predicted values, observe that

$$\sum_{i,g} \hat{y}_{i,g} = \sum_g n_g \mathbf{x}_g^T \hat{\mathbf{a}} = \sum_g n_g \underbrace{\mathbf{x}_g^T \hat{\mathbf{b}}}_{\hat{y}_g} = \sum_g \hat{y}_g \tag{6}$$

Hence, the sum of predictions obtained from Model (1) is the same as the sum of predictions obtained from Model (2), even if partial sums are considered.

□

In the proposition above, the equality should be understood as the equality between estimators. Hence we have the following corollary.

Corollary 2. We define the following matrices

$$\begin{aligned} \mathbf{Y}_g &= [\mathbf{Y}_{1,g} \ \dots \ \mathbf{Y}_{n_g,g}]^T & \mathbf{Y} &= [\mathbf{Y}_1^T \ \dots \ \mathbf{Y}_m^T]^T & \bar{\mathbf{Y}} &= [\bar{Y}_1 \ \dots \ \bar{Y}_m]^T \\ \boldsymbol{\varepsilon}_g &= [\boldsymbol{\varepsilon}_{1,g} \ \dots \ \boldsymbol{\varepsilon}_{n_g,g}]^T & \boldsymbol{\varepsilon} &= [\boldsymbol{\varepsilon}_1 \ \dots \ \boldsymbol{\varepsilon}_m]^T & \mathbf{e} &= [e_1 \ \dots \ e_m]^T \\ \mathbf{x}_g^{n_g} &= \underbrace{[\mathbf{x}_g \ \dots \ \mathbf{x}_g]}_{n_g \text{ times}} & \mathbf{x} &= [\mathbf{x}_1^{n_1} \ \dots \ \mathbf{x}_m^{n_m}] & \bar{\mathbf{x}} &= [\mathbf{x}_1^1 \ \dots \ \mathbf{x}_m^1] \end{aligned}$$

the $(1 \times n_g)$ vectors $\mathbf{1}_{n_g} = [1 \ \dots \ 1]$ and $\mathbf{0}_{n_g} = [0 \ \dots \ 0]$, and the matrix

$$\mathbf{1} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_2} & \dots & \mathbf{0}_{n_m} \\ \mathbf{0}_{n_1} & \mathbf{1}_{n_2} & \dots & \mathbf{0}_{n_m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \dots & \mathbf{1}_{n_m} \end{bmatrix}$$

The OLS estimators are given by

$$\begin{aligned} \hat{\mathbf{A}}_{OLS} &= \operatorname{argmin}_a \{ (\mathbf{Y} - \mathbf{x}^T \mathbf{a})^T (\mathbf{Y} - \mathbf{x}^T \mathbf{a}) \} \\ &= (\mathbf{x} \mathbf{x}^T)^{-1} \mathbf{x} \mathbf{Y} \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{B}}_{OLS} &= \underset{\mathbf{b}}{\operatorname{argmin}} \{ ((\sqrt{\mathbf{11}^T})\bar{\mathbf{Y}} - \sqrt{\mathbf{11}^T}\bar{\mathbf{x}}^T\mathbf{b})^T ((\sqrt{\mathbf{11}^T})\bar{\mathbf{Y}} - (\sqrt{\mathbf{11}^T})\bar{\mathbf{x}}^T\mathbf{b}) \} \\ &= (\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{Y}} \end{aligned}$$

Model (1) on a micro level and Model (2) on a macro level are equivalent, in the sense that

- (i) $\mathbb{E}[\hat{\mathbf{A}}_{OLS}] = \mathbb{E}[\hat{\mathbf{B}}_{OLS}]$ and $\operatorname{Var}[\hat{\mathbf{A}}_{OLS}] = \operatorname{Var}[\hat{\mathbf{B}}_{OLS}]$, when weights n_g are used in Model (2); and
- (ii) $\mathbb{E}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \mathbb{E}\left[\sum_g \hat{Y}_g\right]$ and $\operatorname{Var}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \operatorname{Var}\left[\sum_g \hat{Y}_g\right]$.

Proof. Straightforward calculations lead to $(\mathbf{11}^T)^{-1}\mathbf{1Y} = \bar{\mathbf{Y}}$ and $\bar{\mathbf{x}}\mathbf{1} = \mathbf{x}$.

(i) Let

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{B}}_{OLS}] &= (\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T\mathbb{E}[\mathbf{Y}] \\ &= (\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T(\mathbf{11}^T)^{-1}\mathbf{1}\mathbb{E}[\mathbf{Y}] \\ &= (\mathbf{x}\mathbf{x}^T)^{-1} \mathbf{x}\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\hat{\mathbf{A}}_{OLS}] \end{aligned}$$

For the equality of variances, we have

$$\begin{aligned} \operatorname{Var}[\hat{\mathbf{B}}_{OLS}] &= (\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T\operatorname{Var}[\mathbf{Y}] \left((\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T \right)^T \\ &= (\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T(\mathbf{11}^T)^{-1}\mathbf{1}\operatorname{Var}[\mathbf{Y}]\mathbf{1}^T \left((\mathbf{11}^T)^{-1} \right)^T \left((\bar{\mathbf{x}}\mathbf{11}^T\bar{\mathbf{x}}^T)^{-1} \bar{\mathbf{x}}\mathbf{11}^T \right)^T \\ &= (\mathbf{x}\mathbf{x}^T)^{-1}\mathbf{x}\operatorname{Var}[\mathbf{Y}]\mathbf{x}^T \left((\mathbf{x}\mathbf{x}^T)^{-1} \right)^T \\ &= \operatorname{Var}[\hat{\mathbf{A}}_{OLS}] \end{aligned}$$

(ii) Let

$$\begin{aligned} \mathbb{E}\left[\sum_g \hat{Y}_g\right] &= \mathbb{E}\left[\mathbf{1}_m\mathbf{11}^T\hat{\mathbf{Y}}\right] = \mathbb{E}\left[\mathbf{1}_m\mathbf{11}^T\bar{\mathbf{x}}^T\hat{\mathbf{B}}\right] \\ &= \mathbb{E}\left[\mathbf{1}_m\mathbf{11}^T\bar{\mathbf{x}}^T\hat{\mathbf{A}}\right] = \mathbb{E}\left[\mathbf{1}_n\mathbf{x}^T\hat{\mathbf{A}}\right] \\ &= \mathbb{E}\left[\mathbf{1}_n\hat{\mathbf{Y}}\right] = \mathbb{E}\left[\sum_{i,g} Y_{i,g}\right] \end{aligned}$$

The proof of the equality of variances is similar.

□

2.2. The Quasi-Poisson Regression

A similar result can be obtained in the context of Poisson regressions. A generalized linear model [23] is made up of a linear predictor $\mathbf{x}^T\mathbf{b}$, a link function that describes how the expected value depends on this linear predictor and a variance function that describes how the variance depends on the expected value $\operatorname{Var}[Y] = \varphi V(\mathbb{E}[Y])$, where φ denotes the dispersion parameter. For the Poisson model, the variance is equal to the mean, i.e., $\varphi = 1$ and $V(\mathbb{E}[Y]) = \mathbb{E}[Y]$. This may be too restrictive for many actuarial illustrations, which often show more variation than given by expected values.

We use the term *over-dispersed* for a model where the variance exceeds the expected value. A common way to deal with over-dispersion is a *quasi-likelihood* approach (see [23] for further discussion) where a model is characterized by its first two moments.

Consider either a Poisson regression model, or a quasi-Poisson one,

$$Y_{i,g} \sim \mathcal{P}(\lambda_{i,g}) \text{ or } Y_{i,g} \sim q\mathcal{P}(\lambda_{i,g}) \tag{7}$$

In the case of a Poisson regression,

$$\mathbb{E} [Y_{i,g}] = \lambda_{i,g} = \exp[\mathbf{x}_g^T \mathbf{a} + \ln(1/n_g)] \text{ and } \text{Var} [Y_{i,g}] = \lambda_{i,g} \tag{8}$$

and in the context of a quasi-Poisson regression,

$$\mathbb{E} [Y_{i,g}] = \lambda_{i,g} = \exp[\mathbf{x}_g^T \mathbf{a} + \ln(1/n_g)] \text{ and } \text{Var} [Y_{i,g}] = \varphi_{\text{micro}} \lambda_{i,g} \tag{9}$$

with $\varphi_{\text{micro}} > 0$ for a quasi-Poisson regression ($\varphi_{\text{micro}} > 1$ for over-dispersion). Here again, stacking observations within a cluster yield the following model (on the sum and not the average value, to have a valid interpretation with a Poisson distribution)

$$Y_g = \sum_i Y_{i,g} \sim \mathcal{P}(\lambda_g) \text{ or } Y_g \sim q\mathcal{P}(\lambda_g) \tag{10}$$

In the context of a Poisson regression,

$$\mathbb{E} [Y_g] = \lambda_g = \exp[\mathbf{x}_g^T \mathbf{b}] \text{ and } \text{Var} [Y_g] = \lambda_g$$

and in the context of a quasi-Poisson regression,

$$\mathbb{E} [Y_g] = \lambda_g = \exp[\mathbf{x}_g^T \mathbf{b}] \text{ and } \text{Var} [Y_g] = \varphi_{\text{macro}} \lambda_g \tag{11}$$

with $\varphi_{\text{macro}} > 0$ for a quasi-Poisson regression. Here again, those two models (“micro” and “macro”) are equivalent, in the sense that the following proposition holds.

Proposition 3. *Model (7) on a micro level and Model (10) on a macro level are equivalent in the sense that*

- (i) $\hat{\mathbf{a}}_{MLE} = \hat{\mathbf{b}}_{MLE}$; and
- (ii) $\sum_{i,g} \hat{y}_{i,g} = \sum_g \hat{y}_g$.

Proof. (i) Maximum likelihood estimator of \mathbf{a} is the solution of

$$\sum_{i,g} \left(\frac{y_{i,g} - \exp[\mathbf{x}_g^T \mathbf{a}]}{\varphi_{\text{micro}}} \right) \mathbf{x}_g = \mathbf{0}$$

or equivalently

$$\sum_{i,g} (y_{i,g} - \exp[\mathbf{x}_g^T \mathbf{a}]) \mathbf{x}_g = \mathbf{0}$$

With offsets $\lambda_g^* = \exp[\mathbf{x}_g^T \mathbf{b} + \log(n_g)]$, $g = 1, \dots, m$, maximum likelihood estimator of \mathbf{b} is the solution (as previously, we can remove φ_{macro}) of

$$\sum_g (y_g - n_g \exp[\mathbf{x}_g^T \mathbf{b}]) \mathbf{x}_g = \mathbf{0}$$

$$\sum_{i,g} (y_{i,g} - \exp[x_g^T \mathbf{b}]) \mathbf{x}_g = 0$$

Hence, $\hat{\mathbf{a}} = \hat{\mathbf{b}}$, as (unique) solutions of the same system of equations.

(ii) The sum of predicted values is

$$\begin{aligned} \sum_{i,g} \hat{y}_{i,g} &= \sum_g n_g \hat{\lambda}_{i,g} = \sum_g n_g \exp[x_g^T \hat{\mathbf{a}}] = \sum_g n_g \exp[x_g^T \hat{\mathbf{b}}] \\ &= \sum_g \exp[x_g^T \hat{\mathbf{b}} + \log(n_g)] = \sum_g \hat{\lambda}_g^* = \sum_g \hat{y}_g \end{aligned}$$

□

Nevertheless, as we will see later on, the Corollary obtained in the context of a Gaussian linear model does not hold in the context of a quasi-Poisson regression.

Corollary 4. Model (7) on a micro level and Model (10) on a macro level are asymptotically equivalent for Poisson regressions, in the sense that

- (i) $\mathbb{E}[\hat{\mathbf{A}}_{MLE}] = \mathbb{E}[\hat{\mathbf{B}}_{MLE}]$ and $\text{Var}[\hat{\mathbf{A}}_{MLE}] = \text{Var}[\hat{\mathbf{B}}_{MLE}]$, when n goes to infinity; and
- (ii) $\mathbb{E}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \mathbb{E}\left[\sum_g \hat{Y}_g\right]$ and $\text{Var}\left[\sum_{i,g} \hat{Y}_{i,g}\right] = \text{Var}\left[\sum_g \hat{Y}_g\right]$, when n goes to infinity.

Proof. (i) A classical result of asymptotic theory for maximum likelihood estimators indicates that, under mild regularity conditions, $\mathbb{E}[\hat{\mathbf{A}}_{MLE}] \rightarrow \mathbf{a}$ and $\mathbb{E}[\hat{\mathbf{B}}_{MLE}] \rightarrow \mathbf{b}$ as $n \rightarrow \infty$. Since $\mathbf{a} = \mathbf{b}$, we have $\mathbb{E}[\hat{\mathbf{B}}_{MLE}] = \mathbb{E}[\hat{\mathbf{A}}_{MLE}]$ when $n \rightarrow \infty$. For Model (7), the Fisher information matrix is $\mathbf{I}(\mathbf{A}) = \mathbf{xWx}^T$ and, when $n \rightarrow \infty$, $\text{Var}[\hat{\mathbf{A}}] \rightarrow (\mathbf{xWx}^T)^{-1}$, where $\mathbf{W} = \text{diag}((\lambda_1/n_1)\mathbf{1}_{n_1}, \dots, (\lambda_m/n_m)\mathbf{1}_{n_m})$. For Model (10), we have $\mathbf{I}(\mathbf{B}) = \bar{\mathbf{x}}\mathbf{1W1}^T\bar{\mathbf{x}}^T = \mathbf{xWx}^T$ and, when $n \rightarrow \infty$, $\text{Var}[\hat{\mathbf{B}}] \rightarrow (\mathbf{xWx}^T)^{-1}$.

(ii) By using a similar argument, we have when n goes to infinity

$$\begin{aligned} \mathbb{E}\left[\sum_g \hat{Y}_g\right] &= \mathbb{E}\left[\mathbf{1}_m \mathbf{11}^T \hat{\mathbf{Y}}\right] = \mathbf{1}_m \mathbf{11}^T M_{\hat{\mathbf{B}}}(\bar{\mathbf{x}}^T) \\ &= \mathbf{1}_m \mathbf{11}^T M_{\hat{\mathbf{A}}}(\bar{\mathbf{x}}^T) = \mathbb{E}\left[\mathbf{1}_n \mathbf{1}^T e^{\bar{\mathbf{x}}^T \hat{\mathbf{A}}}\right] = \mathbb{E}\left[\mathbf{1}_n e^{\mathbf{x}^T \hat{\mathbf{A}}}\right] \\ &= \mathbb{E}\left[\mathbf{1}_n \hat{\mathbf{Y}}\right] = \mathbb{E}\left[\sum_{i,g} \hat{Y}_{g,i}\right] \end{aligned}$$

□

In small or moderate-sized samples, it should be noted that $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ may be biased for \mathbf{A} and \mathbf{B} , respectively. Generally, this bias is negligible compared with the standard errors (see [24,25]).

In the quasi-Poisson micro-level model (from Model (7)), as discussed above, the estimator of \mathbf{a} is the solution of the quasi-score function

$$\sum_{i,g} \left(\frac{y_{i,g} - \lambda_{i,g}}{\varphi_{micro}} \right) \mathbf{x}_g = 0$$

which implies $\widehat{\mathbf{a}}_{QLE} = \widehat{\mathbf{a}}_{MLE}$. The classical Pearson estimator for the dispersion parameter φ_{micro} is

$$\widehat{\varphi}_{micro} = \sum_{i,g} \frac{(y_{i,g} - \widehat{y}_{i,g})^2 / \widehat{y}_{i,g}}{\sum_g n_g - (k + 1)} = \sum_{i,g} \frac{(n_g y_{i,g} - \widehat{y}_g)^2 / n_g \widehat{y}_g}{\sum_g n_g - (k + 1)}$$

Empirical evidence (see [26]) support the use of the Pearson estimator for estimating φ because it is the most robust against the distributional assumption. In a similar way, the quasi-Poisson macro-level model (from Model (10)), the estimator of \mathbf{b} is the solution of

$$\sum_g \left(\frac{y_g - \lambda_g}{\varphi_{macro}} \right) \mathbf{x}_g = 0$$

which implies here also $\widehat{\mathbf{b}}_{QLE} = \widehat{\mathbf{b}}_{MLE}$. The dispersion parameter φ is estimated by

$$\widehat{\varphi}_{macro} = \sum_g \frac{(y_g - \widehat{y}_g)^2 / \widehat{y}_g}{m - (k + 1)}$$

Clearly, $\widehat{\varphi}_{micro} \neq \widehat{\varphi}_{macro}$ involving the following results.

Corollary 5. *Model (7) on a micro level and Model (10) on a macro level are not asymptotically equivalent for quasi-Poisson regressions, in the sense that*

- (i) $\mathbb{E}[\widehat{\mathbf{A}}_{QLE}] = \mathbb{E}[\widehat{\mathbf{B}}_{QLE}]$ but $\text{Var}[\widehat{\mathbf{A}}_{QLE}] \neq \text{Var}[\widehat{\mathbf{B}}_{QLE}]$, when n goes to infinity; and
- (ii) $\mathbb{E}\left[\sum_{i,g} \widehat{Y}_{i,g}\right] = \mathbb{E}\left[\sum_g \widehat{Y}_g\right]$ but $\text{Var}\left[\sum_{i,g} \widehat{Y}_{i,g}\right] \neq \text{Var}\left[\sum_g \widehat{Y}_g\right]$, when n goes to infinity.

Proof. (i) The property that variances are not equal is a direct consequence of classical results from the theory of generalized linear models (see [23]), since the covariance matrices of estimators are given by

$$\text{Var} [\widehat{\mathbf{B}}] \rightarrow \widehat{\varphi}_{macro} \left(\mathbf{xWx}^\top \right)^{-1}$$

and

$$\text{Var} [\widehat{\mathbf{A}}] \rightarrow \widehat{\varphi}_{micro} \left(\mathbf{xWx}^\top \right)^{-1} \tag{12}$$

when n goes to infinity. Thus, covariance matrices of estimators are asymptotically equal for the Poisson regression model but differ for the quasi-Poisson model because $\widehat{\varphi}_{micro} \neq \widehat{\varphi}_{macro}$.

- (ii) Since the MLE and the QLE share the same asymptotic distribution (see [23]), the proof is similar to Corollary 4(ii).

□

2.3. Poisson Regression with Random Effect

In the micro-level model described by Equation (7), observations made for the same event (subject) at different periods are supposed to be independent. Within-subject correlation can be included in the model by adding random, or subject-specific, effects in the linear predictor. In the Poisson regression model with random intercept, the between-subject variation is modeled by a random intercept γ which represents the combined effects of all omitted covariates.

Let $Y_g^{(t)}$ represent the sum of all observations from subject t , in the cluster g and

$$\begin{aligned} \mathbb{E} \left[Y_g^{(t)} | \gamma_t \right] &= \exp[\mathbf{x}_g^T \mathbf{a} + \gamma_t] \\ \text{Var} \left[Y_g^{(t)} | \gamma_t \right] &= \mathbb{E} \left[Y_g^{(t)} | \gamma_t \right] \\ \boldsymbol{\gamma} &= \begin{bmatrix} \gamma_1 \\ \dots \\ \gamma_T \end{bmatrix} \sim N_T \left(\mathbf{0}, \sigma^2 \mathbf{I} \right) \end{aligned}$$

where \mathbf{I} is the $(T \times T)$ identity matrix, and $N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the T -dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Straightforward calculations lead to

$$\begin{aligned} \mathbb{E} \left[Y_g^{(t)} \right] &= \exp[\mathbf{x}_g^T \mathbf{a} + \sigma^2/2] \\ \text{Var} \left[Y_g^{(t)} \right] &= \mathbb{E} \left[Y_g^{(t)} \right] \left(1 + \mathbb{E} \left[Y_g^{(t)} \right] \left(\exp[\sigma^2] - 1 \right) \right) \end{aligned}$$

Hence,

$$\text{Var} \left[Y_g^{(t)} \right] > \mathbb{E} \left[Y_g^{(t)} \right], \quad \sigma^2 > 0 \tag{13}$$

This last equation shows that the Poisson regression model with random intercept leads to an over-dispersed marginal distribution for the variable $Y_g^{(t)}$. The maximum likelihood estimation for parameters requires Laplace approximation and numerical integration (see the Chapter 4 of [27] for more details). This model is a special case of multilevel Poisson regression model and estimation can be performed with various statistical softwares such as HLM, SAS, Stata and R (with package lme4).

One may be interested to verify the need of a source of between-subject variation. Statistically, it is equivalent to testing the variance of γ to be zero. In this particular case, the null hypothesis places σ^2 on the boundary of the model parameter space which complicates the evaluation of the asymptotic distribution of the classical likelihood ratio test (LRT) statistic. From the very general result of [28], it can be demonstrated (see [29]) that the asymptotic null distribution of the LRT statistic is a 50/50 mixture of χ_0^2 and χ_1^2 as $\sum_g n_g \rightarrow \infty$. In this case, obtaining an equivalent macro-level model is of little practical interest since the construction of the variance-covariance matrix would require knowledge of the individual (“micro”) data.

3. Clustering and Loss Reserving Models

A loss reserving macro-level model is constructed from data summarized in a table called run-off triangle. Aggregation is performed by occurrence and development periods (typically years). For occurrence period $i, i = 1, 2, \dots, I$, and for development period $j, j = 1, 2, \dots, I$, let $C_{i,j}$ and $Y_{i,j}$ represent the total cumulative paid amount and the incremental paid amount, respectively with $Y_{i,j} = C_{i,j} - C_{i,j-1}, i = 1, \dots, I, j = 2, \dots, I$.

$$\begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,I-1} & C_{1,I} \\ C_{2,1} & C_{2,2} & \dots & C_{2,I-1} & \\ \vdots & \vdots & \ddots & & \\ C_{I,1} & & & & \end{bmatrix}$$

$$\begin{bmatrix} Y_{1,1} = C_{1,1} & Y_{1,2} & \dots & Y_{1,I-1} & Y_{1,I} \\ Y_{2,1} = C_{2,1} & Y_{2,2} & \dots & Y_{2,I-1} & \\ \vdots & \vdots & \ddots & & \\ Y_{I,1} = C_{I,1} & & & & \end{bmatrix}$$

where columns, rows and diagonals represent development, occurrence and calendar periods, respectively. Each incremental cell $Y_{i,j}$ can be seen as a cluster stacking $n_{i,j}$ amounts paid in the same development period j for the occurrence period i . These payments come from M claims and let $Y_{i,j}^{(k)}$ represent the sum of all observations from claims k in the cluster (i, j) . It should be noted that all claims are not necessarily represented in each of the clusters.

To calculate a *best estimate* for the reserve, the lower part of the triangle must be predicted and the total reserve amount is

$$\widehat{R} = \sum_{t=2}^I \widehat{C}_{t,I} - \sum_{t=2}^I C_{t,I-t+1} = \sum_{t=2}^I \sum_{s=I+2-t}^I \widehat{Y}_{t,s}$$

To quantify uncertainty in estimated claims reserve, we consider the mean square error of prediction (MSEP). Let \widehat{R} be a \mathcal{Y} -measurable estimator for $\mathbb{E}[R|\mathcal{Y}]$ and a \mathcal{Y} -measurable predictor for R where \mathcal{Y} represents the set of observed clusters. The MSEP is

$$\begin{aligned} \text{MSEP}_{R|\mathcal{Y}}(\widehat{R}) &= \mathbb{E} \left[\left(\widehat{R} - R \right)^2 | \mathcal{Y} \right] \\ &= \text{Var} [R|\mathcal{Y}] + \left(\widehat{R} - \mathbb{E} [R|\mathcal{Y}] \right)^2 \end{aligned}$$

Independence between R and \mathcal{Y} is assumed, so the equation is simplified as follows

$$\text{MSEP}_{R|\mathcal{Y}}(\widehat{R}) = \text{Var} [R] + \left(\widehat{R} - \mathbb{E} [R] \right)^2$$

and the unconditional MSEP is

$$\text{MSEP}_R(\widehat{R}) = \text{Var} [R] + \mathbb{E} \left[\left(\widehat{R} - \mathbb{E} [R] \right)^2 \right]$$

3.1. The Quasi-Poisson Model for Reserves

3.1.1. Construction

From the theory presented in Subsection 2.2, we construct quasi-Poisson macro- and micro-level models for reserves. For both models, constitutive elements are defined in Table 1.

Table 1. Quasi-Poisson macro- and micro-level models for reserve $(i, j = 1, \dots, I)$. All clusters and all payments are independent.

Components	Macro	Micro
Exp. value	$\mathbb{E} [Y_{i,j}] = \lambda_{i,j}$	$\mathbb{E} [Y_{i,j}^{(k)}] = \lambda_{i,j}$
Inv. link func.	$\lambda_{i,j} = \exp[\mathbf{x}_{i,j}^T \mathbf{b}]$ $= \exp[b_i + b_{I+j}]$ with $b_{I+1} = 0$	$\lambda_{i,j} = \exp[\mathbf{x}_{i,j}^T \mathbf{a} + \log(1/n_{i,j})]$ $= \exp[a_i + a_{I+j} + \log(1/n_{i,j})]$ with $a_{I+1} = 0$
Variance	$\text{Var} [Y_{i,j}] = \varphi_{macro} \lambda_{i,j}$	$\text{Var} [Y_{i,j}^{(k)}] = \varphi_{micro} \lambda_{i,j}$
Pred. value	$\widehat{Y}_{i,j} = \exp[\widehat{b}_i + \widehat{b}_{I+j}]$	$\widehat{Y}_{i,j}^{(k)} = \exp[\widehat{a}_i + \widehat{a}_{I+j} + \log(1/n_{i,j})]$
Known values	\mathcal{Y}_{macro}	\mathcal{Y}_{micro}

As a direct consequence of Proposition 3, the best estimate for the total reserve amount is

$$\widehat{R} = \sum_{(i,j) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \widehat{Y}_{i,j}^{(k)} = \sum_{(i,j) \in \mathcal{K}} \widehat{Y}_{i,j}$$

where \mathcal{K} represents unobserved clusters. For both models, the Proposition 6 gives results for the unconditional MSEF.

Proposition 6. *In the quasi-Poisson macro-level model, the unconditional MSEF is given by*

$$\begin{aligned} \widehat{MSEF}_R(\widehat{R}) &\approx \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{macro} \widehat{Y}_{i,j} \\ &+ \sum_{(i,j), (n,m) \in \mathcal{K}} \widehat{\varphi}_{macro} \widehat{Y}_{i,j} \widehat{Y}_{m,n} \mathbf{x}_{i,j}^T (\mathbf{xWx}^T)^{-1} \mathbf{x}_{n,m} \end{aligned}$$

where \mathbf{x} and \mathbf{W} are defined by Equation (12). The unconditional MSEF for the quasi-Poisson micro-level model is similar with $\widehat{\varphi}_{macro}$ replaced by $\widehat{\varphi}_{micro}$.

Proof. The proof for the macro-level model is done in [25]. For the micro-level model, we have

$$\begin{aligned} MSEF_R(\widehat{R}) &= \text{Var} [R] + \mathbb{E} \left[\left(\widehat{R} - \mathbb{E} [R] \right)^2 \right] \\ &= \sum_{(i,j) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}} + \log(1/n_{i,j})] \\ &+ \sum_{(i,j) \in \mathcal{K}} \sum_{(m,n) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \sum_{t=1}^{n_{m,n}} \text{Cov} \left[\widehat{Y}_{i,j}^{(k)}, \widehat{Y}_{m,n}^{(t)} \right] \\ &= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}}] \\ &+ \sum_{(i,j) \in \mathcal{K}} \sum_{(m,n) \in \mathcal{K}} \sum_{k=1}^{n_{i,j}} \sum_{t=1}^{n_{m,n}} \exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}} + \log(1/n_{i,j})] \exp[\mathbf{x}_{m,n}^T \widehat{\mathbf{a}} + \log(1/n_{m,n})] \\ &\times \text{Cov} \left[\exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}} - \mathbf{x}_{i,j}^T \widehat{\mathbf{a}}], \exp[\mathbf{x}_{m,n}^T \widehat{\mathbf{a}} - \mathbf{x}_{m,n}^T \widehat{\mathbf{a}}] \right] \end{aligned}$$

Although $\widehat{Y}_{i,j}^{(k)}$ is not an unbiased estimator of $\mathbb{E} \left[Y_{i,j}^{(k)} \right]$, the bias is generally of small order and by using the approximation $\exp[x] \approx 1 + x$ for $x \approx 0$, we obtain

$$\begin{aligned} &= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}}] \\ &+ \sum_{(i,j), (m,n) \in \mathcal{K}} \exp[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}} + \mathbf{x}_{m,n}^T \widehat{\mathbf{a}}] \text{Cov} \left[\mathbf{x}_{i,j}^T \widehat{\mathbf{a}}, \mathbf{x}_{m,n}^T \widehat{\mathbf{a}} \right]. \end{aligned}$$

By using the fact that $\widehat{\mathbf{b}} = \widehat{\mathbf{a}}$ and the remark at the end of Subsection 2.2, we obtain

$$= \sum_{(i,j) \in \mathcal{K}} \widehat{\varphi}_{micro} \widehat{Y}_{i,j} + \sum_{(i,j), (m,n) \in \mathcal{K}} \widehat{\varphi}_{micro} \widehat{Y}_{i,j} \widehat{Y}_{m,n} \mathbf{x}_{i,j}^T (\mathbf{xWx}^T)^{-1} \mathbf{x}_{m,n}$$

□

Thus, the difference between the variability in macro- and micro-level models results from the difference between dispersion parameters. Define standardized residuals for both models

$$r_{i,g} = \frac{(y_{i,g} - \hat{y}_{i,g})}{\sqrt{\hat{y}_{i,g}}} \text{ and } r_g = \frac{(y_g - \hat{y}_g)}{\sqrt{\hat{y}_g}}$$

Direct calculations lead to

$$\Psi = \frac{\sum_{i,g} r_{i,g}^2}{\sum_g r_g^2} \leq \frac{\sum_g n_g - (k + 1)}{m - (k + 1)} \rightarrow \hat{\varphi}_{\text{micro}} \leq \hat{\varphi}_{\text{macro}} \tag{14}$$

Thus, if the total number of payments ($\sum_g n_g$) is greater than the value $\Psi(m - (k + 1)) + k + 1$, then the micro-level Model (7) will lead to a greater precision for the best estimate of the total reserve amount and conversely. Adding one or more covariate(s) at the micro level will decrease the numerator of Ψ and will increase the interest of the micro-level model.

3.1.2. Illustration and Discussion

To illustrate these results, we consider the incremental run-off triangle from UK Motor Non-Comprehensive account (published by [30]) presented in Table 2 where each cell (i, j) , $i + j \leq 7$, is assumed to be a cluster g , *i.e.*, the value Y_g is the sum of n_g independent payments. Simulations and computations were performed in R, using packages ChainLadder and gtools. The final reserve amount obtained from the Mack’s model [2] is \$28,655,773.

Table 2. Incremental run-off triangle for macro-level model (in 000’s).

	1	2	3	4	5	6	7
1	3511	3215	2266	1712	1059	587	340
2	4001	3702	2278	1180	956	629	–
3	4355	3932	1946	1522	1238	–	–
4	4295	3455	2023	1320	–	–	–
5	4150	3747	2320	–	–	–	–
6	5102	4548	–	–	–	–	–
7	6283	–	–	–	–	–	–

Based on the construction detailed in Table 1, we consider 2 macro-level models

Model A: $Y_g \sim \mathcal{P}(\lambda_g)$ $\lambda_g = \exp[\mathbf{x}_g^T \mathbf{a}]$

Model B: $Y_g \sim q\mathcal{P}(\lambda_g)$

and 2 micro-level models

Model C: $Y_{i,g} \sim \mathcal{P}(\lambda_{i,g})$ $\lambda_{i,g} = \exp[\mathbf{x}_g^T \mathbf{a} - \log(n_g)]$

Model D: $Y_{i,g} \sim q\mathcal{P}(\lambda_{i,g})$.

To create micro-level datasets from the "macro" one, we perform the following procedure:

1. simulate the number of payments for each cluster assuming $N_g \sim \mathcal{P}(\theta)$, $g = 1, \dots, m$;
2. for each cluster, simulate a $(n_g \times 1)$ vector of proportions assuming $\boldsymbol{\omega}_g = [\omega_1 \ \dots \ \omega_{n_g}]^T \sim \text{Dirichlet}(\mathbf{1})$, $g = 1, \dots, m$;

3. for each cluster, define

$$\begin{bmatrix} Y_{1,g} \\ \vdots \\ Y_{n_g,g} \end{bmatrix} = [\omega_g Y_g], \quad g = 1, \dots, m$$

4. adjust **Model C** and **Model D**; and

5. calculate the best estimate and the MSEP of the reserve by using Proposition 6.

For each value of θ , we repeat this procedure 1000 times and we calculate the average best estimate and the average MSEP. Results are presented in Table 3 and Figure 1. For Poisson regression (**Model A** and **C**), results are similar, which is consistent with Corollary 4. For micro-level models, convergence of \sqrt{MSEP} towards (11 622) is fast. For quasi-Poisson regression (**Model B** and **D**), expected values are equal and Figure 1 shows \sqrt{MSEP} as a function of the expected total number of payments for the portfolio. Above a certain level, (close to 3400 here), accuracy of the “micro” approach exceeds the “macro”. Again, those results are consistent with Corollary 5. Here, we consider that the expected number of payments by cluster (θ) is constant but it would also be possible to consider a mixture model where $(N_g | \Theta_g) \sim \mathcal{P}(\theta_g)$, $g = 1, \dots, m$, and $\Theta_g \sim \text{Gamma}(\alpha, \beta)$. This modification does not change the conclusions. Finally, a comparison of estimated MSEP for both Poisson and quasi-Poisson models confirms the presence of over-dispersion in the data.

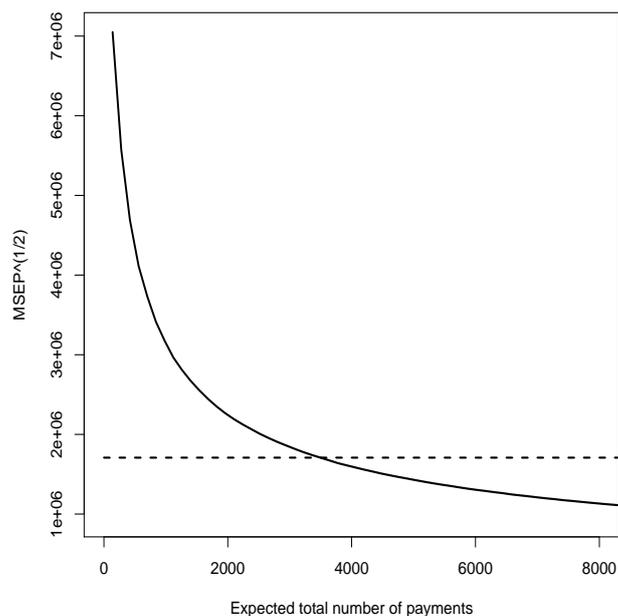


Figure 1. Square root of the mean square error of prediction obtained for **Model D** (solid line) and **Model B** (broken line) from simulated values for increasing expected number of payments for the portfolio.

In order to illustrate the impact of adding a covariate at the micro-level, we define a quasi-Poisson micro-level model with a weakly correlated covariate (**Model E**) and with a strongly correlated covariate (**Model F**). Following a similar procedure, we obtain results presented in the bottom part of Table 3 and in Figure 2.

As opposed to standard classical results on hierarchical models, the average of explanatory variable within a cluster $((1/n_g) \sum_i x_{ig})$ has not been added to the macro-level model (**Model B**), for several reasons,

- (i) impossible to compute that average without individual data;

- (ii) discrete explanatory variables used in the micro-level model; and
- (iii) since claims reserve model have a predictive motivation, it is risky to project the value of an aggregated variable on future clusters.

Table 3. Results.

Method	$\mathbb{E} [\text{Reserve}]$	\sqrt{MSEP}
Mack’s model	28 655 773	1 417 267
Poisson reg.		
Model A	28 655 773	11 622
Model C	28 655 773	11 622
quasi-Poisson reg.		
Model B	28 655 773	1 708 196
Model D	28 655 773	see Figure 1
quasi-Poisson reg.		
Model E ($\rho \approx 0$)	28 657 364	see Figure 2
Model F ($\rho \approx 0.8$)	20 514 566	see Figure 2

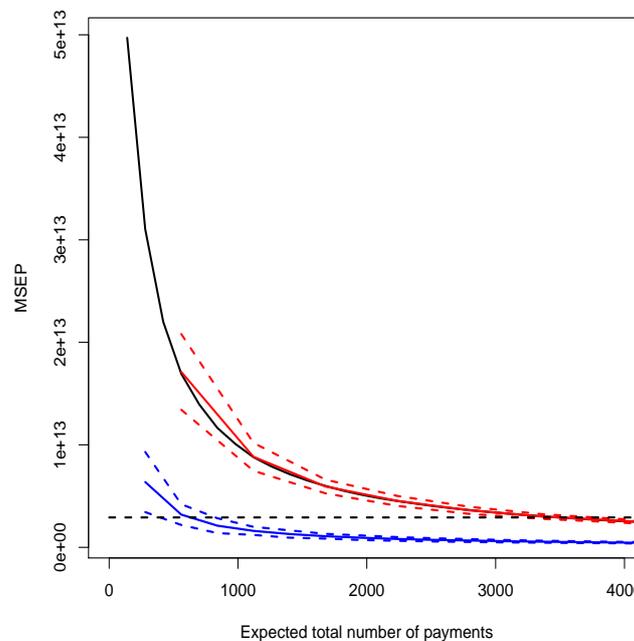


Figure 2. Mean square error of prediction ($\pm 2\sigma$) obtained from simulated values as a function of the expected number of payments for **Model E** (red lines) and **Model F** (blue lines). For comparison purposes, the MSEP obtained for the **Model D** (solid black line) and the **Model B** (broken black line) are added.

With an explanatory variable highly correlated with the response variable, results obtained with **Model D** and **E** are very close. As claimed by Proposition 6 and Equation (14), an explanatory variable highly correlated with the response variable will decrease the value of \sqrt{MSEP} , and lowers the threshold above which the micro-level model is more accurate than the macro-level one.

The quasi-Poisson macro-level model (**Model B**) with maximum likelihood estimators leads to the same reserves as the chain-ladder algorithm and the Mack’s model (see [31]), assuming the clusters exposure, for $(i, j) \in \mathcal{K}$, is one. To obtain similar results with a quasi-Poisson micro-level model (**Model D**), a similar assumption is necessary: exposure of each claim within cluster (i, j) is $1/n_{ij}$. That assumption implies, on a micro level, that predicted individual payments $\hat{Y}_{ij}^{(k)}$ are proportional to $1/n_{ij}$. That assumption has unfortunately no foundation.

In the Poisson and quasi-Poisson micro-level models (**Model C** and **D**), payments related to the same claim, in two different clusters are supposed to be non-correlated. As discussed in the previous Section, it is possible to include dependencies among payments for a given claim using a Poisson regression with random effects.

3.2. The Mixed Poisson Model for Reserves

3.2.1. Construction

From the results obtained in Subsection 2.3, it is possible to construct a micro-model for the reserves that includes a random intercept. The later will allow to model dependence between payments from a given claim. Note that it is not relevant to construct an aggregated model with random effects that could be compared with individual ones. In the context of claims reserves, $Y_g^{(t)}$ represents the sum of payments made for claim t within cluster g . The assumptions of that model (called **model G**) are

$$\begin{aligned} \left(Y_g^{(t)} | \gamma_t \right) &\sim \mathcal{P}(\lambda_g e^{\gamma_t}), & \lambda_g &= \exp[\mathbf{x}_g^T \mathbf{c} + \ln(1/n_g)] \\ \gamma_t &\sim N(0, \sigma^2) \end{aligned}$$

Because of the hierarchical structure of the model, predictions can be derived from several philosophical perspectives (see [32], and references therein). In our example, we focus on the following unconditional predictions

$$\begin{aligned} \left(\hat{Y}_g^{(t)} | \gamma_t \right) &\sim \mathcal{P}(\hat{\lambda}_g e^{\gamma_t}) \\ \hat{\lambda}_g &= \exp[\mathbf{x}_g^T \hat{\mathbf{c}} + \ln(1/n_g)] \\ \text{so that } \mathbb{E} \left[\hat{Y}_g^{(t)} \right] &= \lambda_g e^{\sigma^2/2} \end{aligned}$$

and on the conditional ones, where the unknown component of claim t is predicted by the so-called best linear estimate (that minimizes the MSE) $\tilde{\gamma}_t$ (see [25])

$$\begin{aligned} \left(\tilde{Y}_g^{(t)} | \tilde{\gamma}_t \right) &\sim \mathcal{P}(\hat{\lambda}_g e^{\tilde{\gamma}_t}) \\ \text{so that } \mathbb{E} \left[\tilde{Y}_g^{(t)} \right] &= \lambda_g e^{\tilde{\gamma}_t} \end{aligned}$$

It is then possible to compute the overall best estimate for the total amount of reserves.

3.2.2. Illustration and Discussion

We illustrate these results with the same macro-level dataset. In order to construct a micro-level model from Table 2, we follow a procedure similar to the one described in the previous section,

- 1-3. see previous section;
4. for each accident year, allocate randomly the source (t) of each payment;
5. fit **model G**; and
6. compute the best estimate and the MSE of the reserve.

For a fixed value of θ , the procedure is repeated 1000 times. Various values were considered for θ (10, 25, 50, 100 and 250), and results were similar. In order to avoid heavy tables, only the case where $\theta = 10$ is mentioned here. Simulations and computations were performed with R, relying on package lme4. Final results are reported in Table 4. On Figure 3 we can see predictions of the model on observed data, while on Figure 4 we can see predictions of the model for non-observed cells (with $\pm 2\sigma$ in both cases). For each simulation, a LRT is performed to check the non-nullity of the variance term and to confirm the necessity of including a random intercept in the model, which means that correlation among payments (related to the same claim) is positive. Observe that with the mixed model, the log-likelihood

is approximated using numerical integration, which might bias p -values of the test. To avoid that, p have been confirmed using a bootstrap procedure (using package glmmML).

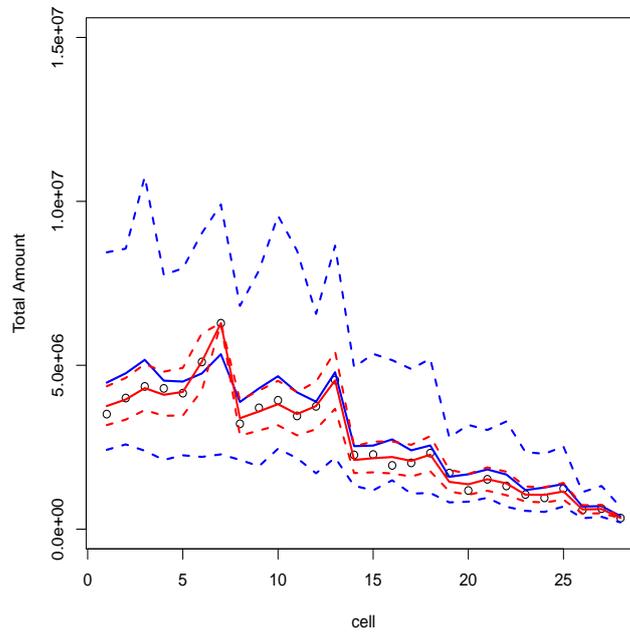


Figure 3. Observed data (circles) with conditional predictions (red lines) and unconditional ones (blue lines) from **Model G** with $\theta = 10$.

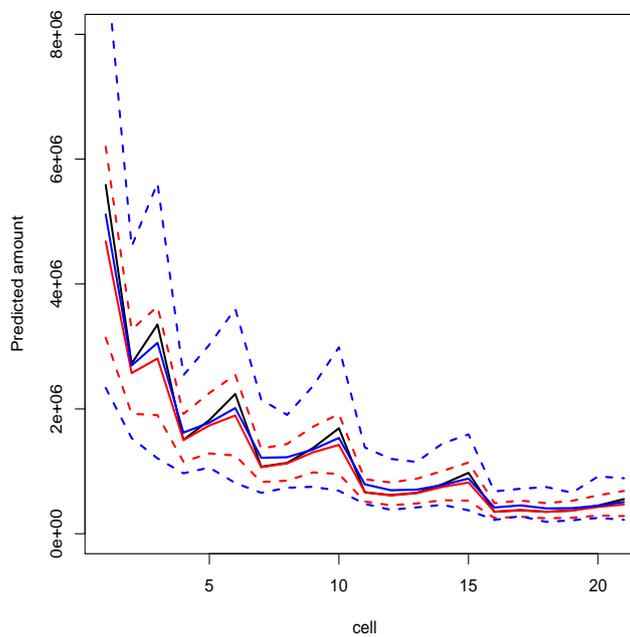


Figure 4. Predictions with the quasi-Poisson macro-level model (strong black line), with conditional predictions (red lines) and unconditional ones (blue lines) from **Model G** with $\theta = 10$.

Table 4. Numerical results for $\theta = 10$. Results for different values of θ are similar.

Modèle	$\mathbb{E} [\text{Reserve}]$	$\sqrt{\text{Var}(\text{Reserve})}$
coll. quasi-Pois.	28 656 423	1 708 216
mixed Poisson non-cond.	27 930 624	3 297 401
mixed Poisson cond.	25 972 947	2 280 902

4. Conclusions

In this article, we study equivalence (as well as non-equivalence) between Poisson and quasi-Poisson regression models, obtained on aggregated (so called macro-level) and non-aggregated (micro-level) datasets, in order to understand when using micro-level models might over-perform macro-level ones. Those models are used here in the context of estimating claims reserves. The uncertainty is quantified using—as in standard macro-level approaches—the MSE. We also investigate the impact of adding micro-level covariates in the model. Finally, we discuss the use of mixed Poisson regression, in the case of micro-level data, that might take into account possible dependence between observations in different clusters.

We illustrate theoretical results on simulated data, generated from cumulated payments, in R. A methodology that allows us to generate such micro-level datasets is described. In a first part, we compare results obtained with Poisson and quasi-Poisson regression, on micro- and macro-level datasets. That comparison reveals that in the context of a quasi-Poisson regression model, the expected number of claims plays a crucial role, since above a given threshold, the micro-level model is more robust than the macro-one. Moreover, the presence (or absence) of covariates can affect this threshold.

In a second part, we analyse predictions obtained using a mixed Poisson regression model, *i.e.*, a Poisson regression model with a random component that characterizes dependence among payments on the same claim, at different dates. The necessity of this random component is verified by using a likelihood-ratio test. That study reveals that such a dependence might have a non-negligible impact on predictions.

Of course, that study is only the first step and several directions for future research can be intuited. For instance, as micro-level models are based on much more observations than macro-level ones, more robust estimation techniques, such as generalized method of moments, can be considered.

Acknowledgments: The authors thank two anonymous referees and the associate editor for useful comments which helped to improve the paper substantially. The first author received financial support from Natural Sciences and Engineering Research Council of Canada (NSERC) and the ACTINFO chair. The second author received financial support from Natural Sciences and Engineering Research Council of Canada (NSERC).

Author Contributions: Both authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Astesan, E. *Les réserves techniques des sociétés d'assurances contre les accidents automobiles*; Librairie générale de droit et de jurisprudence: Paris, France, 1938.
2. Mack, R. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bull.* **1993**, *23*, 213–225.
3. England, P.D.; Verrall, R.J. Stochastic claims reserving in general insurance. *Br. Actuar. J.* **2003**, *8*, 443–518.
4. Van Eeghen, J. Loss reserving methods. In *Surveys of Actuarial Studies 1*; Nationale-Nederlanden: The Hague, The Netherlands, 1981.
5. Taylor, G.C. *Claims Reserving in Non-Life Insurance*; North-Holland: Amsterdam, The Netherlands, 1986.
6. Karlsson, J.E. The expected value of IBNR claims. *Scand. Actuar. J.* **1976**, *1976*, 108–110.
7. Arjas, E. The claims reserving problem in nonlife insurance—Some structural ideas. *ASTIN Bull.* **1989**, *19*, 139–152.
8. Jewell, W.S. Predicting IBNYR events and delays I. Continuous time. *ASTIN Bull.* **1989**, *19*, 25–55.

9. Norberg, R. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bull.* **1993**, *23*, 95–115.
10. Hesselager, O. A Markov model for loss reserving. *ASTIN Bull.* **1994**, *24*, 183–193.
11. Norberg, R. Prediction of outstanding liabilities II: Model variations and extensions. *ASTIN Bull.* **1999**, *29*, 5–25.
12. Hesselager, O.; Verrall, R.J. Reserving in Non-Life Insurance. Available online: <http://onlinelibrary.wiley.com> (accessed on 29 February 2016).
13. Zhao, X.B.; Zhou, X.; Wang J.L. Semiparametric model for prediction of individual claim loss reserving. *Insur. Math. Econ.* **2009**, *45*, 1–8.
14. Zhao, X.; Zhou, X. Applying copula models to individual claim loss reserving methods. *Insur. Math. Econ.* **2010**, *46*, 290–299.
15. Pigeon, M.; Antonio, K.; Denuit M. Individual loss reserving using paid-incurred data. *Insur. Math. Econ.* **2014**, *58*, 121–131.
16. Antonio, K.; Plat, R. Micro-level stochastic loss reserving for general insurance. *Scand. Actuar. J.* **2014**, *2014*, 649–669.
17. Jin, X.; Frees, E.W. Comparing Micro- and Macro-Level Loss Reserving Models. Presentation at ARIA: Madison, WI, USA, 2015.
18. Johansson, A. Claims Reserving on Macro- and Micro-Level. Master's Thesis, Royal Institute of Technology, Stockholm, Sweden, 2015.
19. Friedland, J. *Estimating Unpaid Claims Using Basic Techniques*; Casualty Actuarial Society: Arlington, VA, USA, 2010.
20. Van den Berga, G.J.; van der Klaauw, B. Combining micro and macro unemployment duration data. *J. Econom.* **2001**, *102*, 271–309.
21. Altissimo, F.; Mojon, B.; Zaffaroni, P. *Fast Micro and Slow Macro: Can Aggregation Explain the Persistence of Inflation?* European Central Bank Working Papers, 2007; Volume 0729.
22. Greene, W.H. *Econometric Analysis*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2003.
23. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Chapman & Hall: London, UK, 1989.
24. Cordeiro, G.M.; McCullagh, P. Bias correction in generalized linear models. *J. R. Stat. Soc. B* **1991**, *53*, 629–643.
25. Wüthrich, M.; Merz, M. *Stochastic Claims Reserving Methods*; Wiley Interscience: Hoboken, NJ, USA, 2008.
26. Ruoyan, M. Estimation of Dispersion Parameters in GLMs with and without Random Effects. Stockholm University 2004. Available online: <http://www2.math.su.se/matstat/reports/serieb/2004/rep5/report.pdf> (accessed on 29 February 2016).
27. Snijders, T.A.B.; Bosker, R.J. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*; Sage Publishing: Thousand Oaks, CA, USA, 2012.
28. Self, S.G.; Liang, K.Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **1987**, *82*, 605–610.
29. Dunson, D. *Random Effect and Latent Variable Model Selection*; Lecture Notes in Statistics; Springer-Verlag: New York, NY, USA, 2008; Volume 192.
30. Christofides, S. Regression models based on log-incremental payments. *Claims Reserv. Man.* **1997**, *2*, D5.1–D5.53.
31. Mack, T.; Venter, G. A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insur. Math. Econ.* **2000**, *26*, 101–107.
32. Skrondal, A.; Rabe-Hesketh, S. Prediction in multilevel generalized linear models. *J. R. Stat. Soc. A* **2009**, *172*, 659–687.

