THE IMPACTS OF INDIVIDUAL INFORMATION ON LOSS RESERVING

BY

ZHIGAO WANG, XIANYI WU AND CHUNJUAN QIU

ABSTRACT

The projection of outstanding liabilities caused by incurred losses or claims has played a fundamental role in general insurance operations. Loss reserving methods based on individual losses generally perform better than those based on aggregate losses. This study uses a parametric individual information model taking not only individual losses but also individual information such as age, gender, and so on from policies themselves into account. Based on this model, this study proposes a computation procedure for the projection of the outstanding liabilities, discusses the estimation and statistical properties of the unknown parameters, and explores the asymptotic behaviors of the resulting loss reserving as the portfolio size approaching infinity. Most importantly, this study demonstrates the benefits of individual information on loss reserving. Remarkably, the accuracy gained from individual information is much greater than that from considering individual losses. Therefore, it is highly recommended to use individual information in loss reserving in general insurance.

KEYWORDS

General insurance, individual data, loss reserving, IBNR and RBNS claims, asymptotic analysis.

1. INTRODUCTION

In the operations of insurance companies and other financial institutions, the terms *loss reserve* and *loss reserving* generally refer to the funds kept by the operators. The purpose of these funds is to protect institutions from potential ruins caused by huge claims and payments. In general insurance industry, in particular, adequate assets are required to cover the outstanding liabilities from the written insurance policies. Moreover, insurance industry authorities commonly mandate adequate assets as a fundamental regulatory requirement

Astin Bulletin 51(1), 303-347. doi:10.1017/asb.2020.42 © 2020 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

(e.g., Solvency II Directive put forward by the European Union, which has been in effect since January 1 2016).

For companies operating general insurance, however, it is not an easy task to compute the reserves based on all the observed data for two intertwined reasons. First, the uncertainty about loss distributions requires statistical methodologies to estimate the unknown projections of outstanding liabilities. The second reason stems from two types of inherent lags: the lags between the occurrences of claims events and their reports to the insurers (reporting delays (R-delays)) and the lags of the reported claims between times of the reports and their final settlements (i.e., settlement delays (S-delay)), which depend on various commercial and legal factors. These lags contribute to difficulties in the statistical procedures, though the recent improvements of communication techniques have largely alleviated the delays in reporting claims. The study and practice of loss reserves are generally divided into two areas: incurred but not reported (IBNR) claims and reported but not settled (RBNS) claims, influenced by the R-delays and S-delays of claims, respectively.

The most successive algorithms have so far been the well-known chain-ladder method and its related versions, all of which are based on the so-called aggregate or macro data summarized in run-off triangles. Using this approach, any computation of reserves can be performed via pencil and paper. An extensively accepted distribution-free stochastic understanding of chain-ladder method is proposed by Mack (1993). Other excellent contributions can be found in Gogol (1993) and Verrall (2000) who discussed some parametric settings of the models and England and Verrall (2002, 2006) for credibility, exact Bayesian and generalized linear models, just to name a few. One can also be referred to the monograph by Wüthrich and Merz (2008), and the references therein, for a comprehensive summary of the details of their theories and algorithms.

Another branch of research, stemming from Arjas (1989) and Norberg (1993, 1999), base loss reserves on individual or micro losses. This approach entails modeling the development processes of individual claims using marked Poisson's processes in continuous time, under which the loss reserves can still be computed in closed mathematical forms. The most recent extensions in this branch include Badescu *et al.* (2016) and Yu and He (2016) who modeled the developments using marked Cox processes (also known as double stochastic processes). These processes are more flexible than Poisson's processes, which have equal mean and variance functions and, thus, do not universally fit all claims occurrences.

The main argument supporting the algorithms based on individual data is that, from a statistical perspective, the data employed in the classical run-off triangles are not necessarily sufficient statistics for the unknown parameters. This means significant statistical information in the raw data collected from individuals may be wasted. The past 30 years have seen a series of research efforts on statistical models to compute loss reserves directly from individual losses. In addition to what have been mentioned above, some examples are

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

Jewell (1989, 1990), who used fully parametric Bayesian models under a continuous and discrete time framework to fit the occurrences and their developments of claims; Zhao and Zhou (2010), who considered the R-delays so as to predict the IBNR reserves; and Larsen (2007), who revisited the marked Poisson's processes but with a discrete time setting as a practically feasible skeleton of the model proposed by Norberg (1993).

It appears more convenient and more applicable to practice to statistically model the individual loss reserving with a discrete time framework, similar to Hesselager (1995), Pigeon *et al.* (2013, 2014), and Godecharle and Antonio (2015). Along this line, Huang *et al.* (2015a, b, 2016) demonstrated a significant reduction in the mean squared error of loss reserves using individual/micro data from the popular aggregate data models that typically employ such methods as the chain-ladder and Bornhuetter–Ferguson. They did this by, respectively, deriving their asymptotic variances and numerically comparing them, thus pointing out a promising direction for more informative loss reserving. The fundamental philosophy from statistics maintains that the more the relevant information, the greater the accuracy is.

The main price of employing individual data is the requirement for computational capacity. This is especially the case for general insurance, in which a portfolio generally contains numerous written policies, even reaching a few millions in certain lines of the insurance business. However, with the rapidly increasing in computing power nowadays in terms of central processing units and graphical processing units and techniques such as parallel and distributed computation in huge computer clusters, computation with individual data is no longer a problem, at least for the algorithms that have so far been developed. Recently, some researchers even analyzed the granular models based on weekly and daily recorded data. This was to reflect the heterogeneity in R-delays on the occurrence dates of the claims and the strong weekday and holiday patterns leading to less claims being reported during the weekends and holidays. Meanwhile, Badescu *et al.* (2016) and Avanzi *et al.* (2016) modeled the accident arrival processes at a weekly level using Cox processes. Verrall and Wüthrich (2016), Verbelen *et al.* (2017), and Crevecoeur *et al.* (2019) computed daily reserves based on daily recorded data. Those research efforts have allowed for clearer understanding of the claims development processes. Increasing computing power and refining techniques have supported the newly emerging trend of cooperating machine learning (including deep learning) for loss reserving, as in Gabrielli and Wüthrich (2018), Wüthrich (2018), Kuo (2019), among others.

In recent times, data generated by the insurance community have sharply increased not only in volume and in time granular, but also in structural complexity, owing to rapidly increasing organizational capacities of data collection, collation, and storage. For example, in motor car insurance, one of the most popular strategies for rate making is to incorporate driver behavior, which is recorded by certain intelligent terminals as an aspect of usage-based insurance (a bonus-malus system). Exploring these data's usage can shed light on how to increase the accuracy of outstanding liability projections. In insurance

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

disciplines other than loss reserving, considerable efforts have been made to use individual features , see, for example, Denuit *et al.* (2007) and the references therein. However, while the claims development heterogeneity over time has clearly been recognized, the one over insureds has so far attracted only scarcely few attention in the context of loss reserving. By heterogeneity among insureds, we mean that claims development patterns are influenced by factors that reflect individual heterogeneity among individual policies. For motor car insurance, these factors include, for examples, a policyholder's age, years of driving experience, gender as well as an insured vehicle's price, fuel used, brand class, and geographical location, contract type of the insurance and so on, whereas for health insurance, these factors could include the policyholder's age, gender, health status, and others that have not been recognized by the insurance industries. Fung *et al.* (2020) proposed a class of models of certain desired appealing theoretical properties to analyze IBNR loss reserving. Their models combined the ideas of data transform and distribution mixture to accommodate the heavy-tailed behavior, complex distributional characteristics such as multimodality and peculiar links between policyholders' risk profile and claim amounts exhibited in real datasets. The real data analysis in Fung *et al.* (2020) show a clear evidence for the heterogeneity among insureds in a motor car insurance business. Some more evidences of the heterogeneity can also be found in the real data analysis part of the current paper for a health insurance business (see Section 4.2, especially Figure 4 and Table 5, for details).

This study explores and quantitatively characterizes the possible effects of individual information on loss reserving. It advocates using this approach to loss reserving in the current era of highly developed computing capacity. We are primarily for a theoretical purpose other than to propose a parametric model for loss reserving. Incorporating the individual features into a comprehensive regression model to fit the claims development data in general insurance allow for a more accurate projection of a portfolio's outstanding liabilities. To be specific, this study theoretically demonstrates that, in terms of a particularly specified parametric model for claims development (which nevertheless effectively fits a real data also), the incorporation of insureds' individual information can significantly improve the accuracy of the loss reserve projection. Meanwhile, the reduction in accuracy by incorporating false covariates is relatively limited. Moreover, the advantages of using this approach are more significant than those from considering individual losses. See Theorems 3.4 and 3.5, the subsequent discussion and the simulations in Section 3.4 for more details.

The paper is organized as follows. Section 2 formulates the model and discusses the maximum likelihood estimates (MLEs) of the unknown parameters and the theoretical properties in terms of asymptotic analysis. Based on these theoretical properties, Section 3 derives the corresponding formula of loss reserves and the asymptotic properties of loss reserving, with and without individual information. It also reports a small simulation study in order to exhibit the behaviors of the loss reserving method in fixed size portfolio. Section 4

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

addresses a thorough real data analysis. Section 5 concludes the paper with a few remarks. All proofs are deferred to the Appendix.

2. A CLAIMS DEVELOPMENT MODEL USING INDIVIDUAL INFORMATION

This preliminary section consists of two parts. The first concisely describes an individual claims development. It considers individual information to reflect the individual features of policies and the data structure to loss reserving and, in addition, provides the distributional assumptions. The second part is for the maximum likelihood estimation and its theoretical properties for later use in loss reserving.

2.1. Claims development

In general insurance, a dataset for loss reserving, regardless of whether it is in macro or micro form, is typically organized through periods of fixed length (conventionally referred to as "(accident) years"), as analyzed by, for example, Huang *et al*. (2015a, b) excluding the individual features:

(1) The entire observation horizon is composed of n calendar years, at the end of which, (referred to also a date n), a loss reserving is made.
(2) In a year, say i, there are m_i effective insurance policies (exposures), coded by (i, k), $k = 1, 2, \ldots, m_i$ and associated with every policy (i, k) is a random element:

$$\boldsymbol{E}_{ik} := \left(r_{ik}, \boldsymbol{x}_{ik}; N_{ik}; \{(U_{ikl}, V_{ikl}, Y_{ikl})\}_{l=1}^{N_{ik}}\right), \tag{2.1}$$

consisting of
   (i) a risk exposure $r_{ik} \in (0, 1]$ in that year,
   (ii) a number N_{ik} of incurred claims from this individual, which is possibly incompletely observed at the evaluation date n due to R-delays,
   (iii) accordingly a sequence of chronologically recorded claims data $(U_{ikl}, V_{ikl}, Y_{ikl})$ of R-delays U_{ikl} (the time lags between the occurrences of accidents and their reports to the insurance company), S-delays V_{ikl} (the time lags between their reports to the insurance company and final settlements), and claim amounts Y_{ikl} paid in a lump sum at their settlements, $l = 1, 2, \ldots, N_{ik}$ and
   (iv) especially, an observable d-vector covariates $\boldsymbol{x}_{ik} = (1, x_{ik1}, \ldots, x_{ik,d-1})'$ to indicate the individual information, on which the joint distribution of the frequency of claims, R-delays, S-delays, and payments may depend, where d is a positive integer.
(3) All the individual observations $\boldsymbol{E}_{ik}, k = 1, 2, \ldots, m_i, i = 1, 2, \ldots, n$ are iid from a population $\boldsymbol{E} := (r, \boldsymbol{x}; N; \{(U_l, V_l, Y_l)\}_{l=1}^{N})$, which is the individual observation from a representative policy.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

(4) Because the analyses proceed basically in terms of large sample approximations as $m \to \infty$, where $m = \sum_{i=1}^{n} m_i$ is the total number of policies in the n accident years, we make an assumption that $m_i/m \to \kappa_i$ with constants $\kappa_i, i = 1, 2, \ldots, n$.

Following the conventional terminologies, a claim (U_l, V_l, Y_l) incurred in year i is called settled, RBNS and IBNR if its R- and S-delays (U, V) take values in $\mathscr{A}_i^s = \{(u, v) : i + u + v \le n, \}$ (settled), $\mathscr{A}_i^{rbns} = \{(u, v) : i + u \le n < i + u + v\}$ (RBNS) and $\mathscr{A}_i^{ibnr} = \{(u, v) : i \le n < i + u\}$ (IBNR), respectively.

The joint distribution of the representative claim development $\boldsymbol{E}$ is specified by the product of the marginal distribution of $(r, \boldsymbol{x})$ and the conditional distribution of $(N; \{(U_l, V_l, Y_l)\}_{l=1}^{N})$, as in what follows.

Distribution Assumption 1 (Distribution of $\boldsymbol{E}$). *For any given vector $\boldsymbol{\beta}$, the pair $(r, \boldsymbol{x})$ is arbitrarily distributed with $\mathbb{E}[r \exp(\boldsymbol{x}'\boldsymbol{\beta})] < \infty$. Given $(r, \boldsymbol{x})$, the three random parts N, $\{U_l\}_{l=1}^{N}$, and $\{(V_l, Y_l)\}_{l=1}^{N}$ are mutually independent of each other with each distributed as follows.*

(1) **Claims number** N. *$N \sim Possion(r \exp(\boldsymbol{x}'\boldsymbol{\beta}))$.*

(2) **R-delays** $\{U_l\}_{l=1}^{N}$. *There exists a maximum R-delay D^r such that $\{U_l\}_{l=1}^{N} \overset{iid}{\sim} U$ with*

$$\Pr(U = u|\boldsymbol{x}) := p_u = p_u(\boldsymbol{x}; \boldsymbol{\pi}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\pi}_u)}{\sum_{j=0}^{D^r} \exp(\boldsymbol{x}'\boldsymbol{\pi}_j)}, u = 0, 1, 2, \ldots, D^r, \textit{ with } \boldsymbol{\pi}_0 = \boldsymbol{0}.$$

(3) **Settlements** $\{(V_l, Y_l)\}_{l=1}^{N}$. *Let $\{(V_l, Y_l)\}_{l=1}^{N} \overset{iid}{\sim} (V, Y)$ and there also exists a maximum S-delay D^s such that*

$$\Pr(V = v|\boldsymbol{x}) := q_v = q_v(\boldsymbol{x}; \boldsymbol{\rho}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\rho}_v)}{\sum_{j=0}^{D^s} \exp(\boldsymbol{x}'\boldsymbol{\rho}_j)}, v = 0, 1, \ldots, D^s, \textit{ with } \boldsymbol{\rho}_0 = \boldsymbol{0}.$$

To formulate the distribution of Y given V, introduce augmented vectors $\boldsymbol{x}_v = (\boldsymbol{x}', \boldsymbol{\delta}_v')'$, where $\boldsymbol{\delta}_v$ is a D^s-dimensional vector such that $\boldsymbol{\delta}_0 = \boldsymbol{0}$ and, for $v = 1, 2, \ldots, D^s$, $\boldsymbol{\delta}_v$ is a unit vector with 1 at the v-th position. Then, Y follows a distribution with density $f(y; \eta_v, \sigma), y \in (0, +\infty)$, so that $\mathbb{E}[Y|\boldsymbol{x}, V = v] = \mu_v(\eta_v, \sigma)$ is a function of η_v and σ, where η_v is linked to $\boldsymbol{x}_v'\boldsymbol{\gamma}$ by means of $g(\eta_v) = \boldsymbol{x}_v'\boldsymbol{\gamma}$ for some differentiable increasing function g.

The distribution specification above involves the vectors of unknown parameters:

$$\begin{cases} \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{d-1})' \text{ for claims number } N, \\ \boldsymbol{\pi} = (\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \ldots, \boldsymbol{\pi}_{D^r}')' \text{ for R-delays } U_l \text{ with } \boldsymbol{\pi}_u = (\pi_{u0}, \pi_{u1}, \ldots, \pi_{u,d-1})', u = 1, 2, \ldots, D^r, \\ \boldsymbol{\rho} = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \ldots, \boldsymbol{\rho}_{D^s}')' \text{ for S-delays } V_l \text{ with } \boldsymbol{\rho}_v = (\rho_{v0}, \rho_{v1} \ldots, \rho_{v,d-1})', v = 1, 2, \ldots, D^s, \\ \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{d-1}, \gamma_d, \ldots, \gamma_{d-1+D^s})' \text{ and } \sigma \text{ for claims severity } Y_l, \end{cases} \tag{2.2}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

where $\boldsymbol{\beta}$, $\boldsymbol{\pi}_u$s as well as $\boldsymbol{\rho}_v$s are d-dimensional vectors and $\boldsymbol{\gamma}$ is a $(d + D^s)$-dimensional vector. An abbreviation $\boldsymbol{\theta}$ is sometimes used to represent the whole vector of parameters $(\boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{\rho}', \boldsymbol{\gamma}', \sigma)$.

While it is not the primary purpose of this study to propose a practically useful model, the following remarks explain why loss reserving is discussed under Distribution Assumption 1:

(1) There are basically two cases for the supports of R- and S-delays: finite and infinite. It would be *a priori* known (generally read from the items of the insurance contracts) if the supports are finite or infinite before any loss reserving is taken care of. Even for the case the delays take unrestricted values, if the probability to take values over certain limits is quite small, one can safely assume a capped delays by cutting off the tails with probability small enough. As a result, the assumption of capped R- and S-delays is reasonable in many real insurance businesses, especially for such insurance without very much high claims payments. An example is the general health insurance. The assumption of capped delays has been extensively adopted in such traditional methods as chain-ladder algorithm. If the tails cannot be safely cut off, however, the models such as the one proposed in Fung *et al.* (2020) or some others would be more suitable. From the statistical point of view, for their distributions to be reasonably estimated with observations over a finite number of years, the number of unknown parameters to be estimated must be finite. Here, the former is taken, whereas Crevecoeur *et al.* (2019) and Fung *et al.* (2020), for example, took the latter.

(2) Logistic regression is generally the starting point for analyzing categorical responses. The most recent examples in insurance include, for example, Boj and Costa (2018) and Heras *et al.* (2018). If the coefficients included in $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ are zero, the setting above reduces to the one discussed by Huang *et al.* (2015a, b).

(3) The assumption on the conditional distribution of claims payments subsumes exponential families such as gamma, inverse Gaussian, and log-normal distribution and also includes distributions such as Pareto II that does not belong to exponential families. In our real data analysis, the Pareto II distribution fits the claims payments well.

(4) For convenience, we assume that $\max(D^r, D^s) < n \leq D^r + D^s + 1$. The first inequality is to make $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ to be reasonably estimated and the second inquality does not reduce the generality of the model because for the case $n > D^r + D^s + 1$, one can simply recode the first $n - D^r - D^s$ as year 1.

For easy reference, in the following collected are some necessary notation that will frequently appear in this paper.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

Notation 2.1.

(1) The following will always be used without a claim:
- $\boldsymbol{\delta}_u$ *for unit vectors with* 1 *at component* u *and* $\boldsymbol{\delta}_0 = \mathbf{0}$, *of which dimensions can be read from context,*
- $\mathbf{I}_D$ *the identity matrix of dimension* D,
- $\mathbb{1}_D$ *the vector of dimension* D *with* 1 *everywhere,*
- $\otimes$ *the Kronecker product.*

(2) Write $\boldsymbol{p} = (p_1, p_2, \ldots, p_{D^r})'$, $\log \boldsymbol{p} = (\log p_1, \log p_2, \ldots, \log p_{D^r})'$, *and* $P_u = \sum_{j=0}^{u} p_j$ *and write* $\boldsymbol{q}_v = (q_1, q_2, \ldots, q_v)'$, $\overline{\boldsymbol{q}}_v = (q_{v+1}, q_{v+2}, \ldots, q_{D^s})'$, $\boldsymbol{q} = \boldsymbol{q}_{D^s}$ *and* $\log \boldsymbol{q} = (\log q_1, \log q_2, \ldots, \log q_{D^s})'$, *so that, by the identities* $\frac{\partial p_u}{\partial \boldsymbol{\pi}} = p_u(\boldsymbol{\delta}_u - \boldsymbol{p}) \otimes \boldsymbol{x}$ *and* $\frac{\partial q_v}{\partial \boldsymbol{\rho}} = q_v(\boldsymbol{\delta}_v - \boldsymbol{q}) \otimes \boldsymbol{x}$, *it follows that:*

$$\frac{\partial \boldsymbol{p}'}{\partial \boldsymbol{\pi}} = (\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') \otimes \boldsymbol{x}, \qquad \frac{\partial \boldsymbol{q}'}{\partial \boldsymbol{\rho}} = (\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}') \otimes \boldsymbol{x},$$

$$\frac{\partial \log \boldsymbol{p}'}{\partial \boldsymbol{\pi}} = (\mathbf{I}_{D^r} - \boldsymbol{p}\mathbb{1}'_{D^r}) \otimes \boldsymbol{x}, \qquad \frac{\partial \log \boldsymbol{q}'}{\partial \boldsymbol{\rho}} = (\mathbf{I}_{D^s} - \boldsymbol{q}\mathbb{1}'_{D^s}) \otimes \boldsymbol{x}. \tag{2.3}$$

Further denote by $Q_v = \sum_{t=0}^{v} q_t$ *and* $\bar{Q}_v = \sum_{t=v+1}^{D^s} q_t$, $v = 0, 1, \ldots, D^s - 1$ *the cumulative distribution function and tail probability of the S-delays, respectively, which prove quite helpful in discussing the S-delays subject to censoring. Also write* $\bar{\boldsymbol{Q}} = (\bar{Q}_0, \bar{Q}_2, \ldots, \bar{Q}_{D^s-1})'$. *Thoroughly, we take the convention* $\sum_{j=j_1}^{j_2} \cdot = 0$ *if* $j_1 > j_2$.

(3) Notation $D_i^r = D^r \wedge (n-i)$ *and* $D_i^s = D^s \wedge (n-i)$ *are also used to represent, respectively, the possibly observed maximum R- and S-delays for accident year i.*

2.2. Individual likelihood

To establish the likelihood and estimate the parameters, the likelihood of an individual in every year, say i is described in the following points:

- The potential R-delays $\{U_l : l = 1, 2, \ldots, N\}$ are subject to a common right truncation by $n-i$ so that U_l can be observed only when $U_l < n-i$. Denote by $N^r(\leq N)$ the number of the reported claims from the policy, and with no loss of generality denote by $\{U_l : l = 1, 2, \ldots, N^r\}$ the observed R-delays.
- The potential S-delays $\{V_l : l = 1, 2, \ldots N^r\}$ are subject to a right censoring, such that the observed are random pairs $(Z_l, C_l) := (V_l \wedge (n-i-U_l), C_l)$ with $C_l = 1$ if $V_l \leq n-i-U_l$ (i.e., V_l is observed) and zero otherwise. Use N^s to denote the number of the settled claims out of the N^r reported.
- The payments Y_l are subject to a relevant right truncation in the sense that Y_l is observable only when $C_l = 1$.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

To summarize, the observed part of a claim development $\boldsymbol{E}$ is represented by

$$\boldsymbol{E}^o = \begin{cases} (r, \boldsymbol{x}; N^r, N^s; & \{U_l; (Z_l, C_l)\}_{l=1}^{N^r}; \quad \{Y_l\}_{l=1}^{N^s}), & \text{if } N^r \geq N^s > 0, \\ (r, \boldsymbol{x}; N^r, 0; & \{U_l; (Z_l, C_l)\}_{l=1}^{N^r} \quad * \quad), & \text{if } N^r > 0 \text{ and } N^s = 0, \\ (r, \boldsymbol{x}; 0, \quad 0; & * \quad\quad * \quad), & \text{if } N^r = 0, \end{cases} \tag{2.4}$$

where “*” indicates the missing parts caused by the truncation. Then, under Distribution Assumption 1, the log-likelihood in terms of $\boldsymbol{E}^o$ in year i is

$$l_{\boldsymbol{E}^o}(\boldsymbol{\theta}) = N^r \boldsymbol{x}'\boldsymbol{\beta} - r\exp(\boldsymbol{x}'\boldsymbol{\beta})P_{D_i^r} + \sum_{l=1}^{N^r}\big[\log p_{U_l} + C_l \log q_{V_l} + (1 - C_l)\log \bar{Q}_{n-i-U_l} + C_l \log f(Y_l; \eta_{V_l}, \sigma)\big], \tag{2.5}$$

up to an additive constant (not mentioned below), where D_i^r is defined in Notation 2.1 (3) and $\sum_{l=1}^{0} \cdot = 0$ is assumed to accommodate the case $N^r = 0$.

Split $N^r = \sum_{u=0}^{D_i^r} N_u^r$ into a D^r-vector $\boldsymbol{N}_i^r = (N_1^r, N_2^r, \ldots, N_{D_i^r}^r, 0, \ldots, 0)'$ and $N^s = \sum_{v=0}^{D_i^s} N_v^s$ into a D^s-vector $\boldsymbol{N}_i^s = (N_1^s, N_2^s, \ldots, N_{D_i^s}^s, 0, \ldots, 0)'$ with N_u^r and N_v^s the numbers of reported claims with R-delay u and settled claims with S-delay v, respectively. Further split the total number of RBNS claims $N^{rbns} := N^r - N^s$ into $\boldsymbol{N}_i^{rbns} = (N_0^{rbns}, N_1^{rbns}, \ldots, N_{D_i^r}^{rbns}, 0, \ldots, 0)'$ with N_u^{rbns} as the number of RBNS claims with R-delay u.

Moreover, denote by $Y_{vl}, l = 1, 2, \ldots N_v^s$ the settled claims payments with S-delay v. Then, the log-likelihood can be simplified further to:

$$l_{\boldsymbol{E}^o} = N^r \boldsymbol{x}'\boldsymbol{\beta} - r\exp(\boldsymbol{x}'\boldsymbol{\beta})P_{D_i^r} + (N_0^r, \boldsymbol{N}_i^{r\prime})\big(\log p_0, \log \boldsymbol{p}'\big)' + (N_0^s, \boldsymbol{N}_i^{s\prime})(\log q_0, \log \boldsymbol{q}')' + \sum_{u=0}^{D_i^r} N_u^{rbns} \log \bar{Q}_{n-i-u} + \sum_{v=0}^{D_i^s}\sum_{l=1}^{N_v^s} \log f(Y_{vl}; \eta_v, \sigma). \tag{2.6}$$

Because the terms $1-3$, $4-5$, and 6 here involve only parameters $(\boldsymbol{\beta}, \boldsymbol{\pi})$, $\boldsymbol{\rho}$, and $(\boldsymbol{\gamma}, \sigma)$, respectively, the MLEs of $(\boldsymbol{\beta}, \boldsymbol{\pi})$, $\boldsymbol{\rho}$, and $(\boldsymbol{\gamma}, \sigma)$ can be derived separately and involve only the numbers N_u^r's, (N_v^s, N_u^{rbns})'s, and N_v^s's and their payments, respectively. Note that $N_u^{rbns} = 0$ if $n - i - u \geq D^s$ (i.e., $0 \leq u \leq n - i - D^s$). The summation in the second to the last term of (2.6) can also be represented by another summation $\sum_{u=(n-i-D^s+1)_+}^{D_i^r}$.

Denote by $\boldsymbol{p}_{D_i^r} = (p_1, \ldots, p_{D_i^r}, 0, \ldots, 0)'$ the truncated R-delay probabilities (D^r-dimensional). The following lemma is for individual score functions and information matrix.

Lemma 2.1 (Individual scores and information). *The following holds for an individual in year i:*

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

(1) the score functions are

$$\begin{pmatrix} \frac{\partial l_{E^o}}{\partial \boldsymbol{\beta}} \\ \frac{\partial l_{E^o}}{\partial \boldsymbol{\pi}} \\ \frac{\partial l_{E^o}}{\partial \boldsymbol{\rho}} \\ \frac{\partial l_{E^o}}{\partial (\boldsymbol{\gamma}', \sigma)'} \end{pmatrix} = \begin{pmatrix} (N^r - r\exp(\boldsymbol{x}'\boldsymbol{\beta})P_{D_i^r})\boldsymbol{x} \\ \left[\boldsymbol{N}_i^r - r\exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{p}_{D_i^r} - (N^r - r\exp(\boldsymbol{x}'\boldsymbol{\beta})P_{D_i^r})\boldsymbol{p}\right] \otimes \boldsymbol{x} \\ \left[\boldsymbol{N}_i^s + \sum_{u=0}^{D_i^r} \frac{N_u^{rbns}}{\bar{Q}_{n-i-u}} \begin{pmatrix} \boldsymbol{0} \\ \bar{\boldsymbol{q}}_{n-i-u} \end{pmatrix} - N^r \boldsymbol{q}\right] \otimes \boldsymbol{x} \\ \sum_{v=0}^{D_i^s} \sum_{l=1}^{N_v^s} \frac{\partial \log f(Y_{vl}; \eta_v, \sigma)}{\partial (\boldsymbol{\gamma}', \sigma)'} \end{pmatrix}. \tag{2.7}$$

(2) The Fisher information matrix is the expectation of the block diagonal matrix:

$$I = r\exp(\boldsymbol{x}'\boldsymbol{\beta}) \begin{pmatrix} I_{1i} \otimes \boldsymbol{x}\boldsymbol{x}' & & \\ & I_{2i} \otimes \boldsymbol{x}\boldsymbol{x}' & \\ & & I_{3i} \end{pmatrix},$$

where

$$\begin{aligned} I_{1i} &= \sum_{u=0}^{D_i^r} p_u \begin{pmatrix} 1 \\ \boldsymbol{\delta}_u - \boldsymbol{p} \end{pmatrix} (1, \boldsymbol{\delta}_u' - \boldsymbol{p}'), \\ I_{2i} &= P_{n-i-D^s}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}') + \sum_{v=(n-i-D^r)_+}^{(D^s-1)\wedge(n-i)} p_{n-i-v} \begin{pmatrix} \mathrm{diag}(\boldsymbol{q}_v) - \boldsymbol{q}_v\boldsymbol{q}_v' & -\boldsymbol{q}_v\bar{\boldsymbol{q}}_v' \\ -\bar{\boldsymbol{q}}_v\boldsymbol{q}_v' & \frac{Q_v}{\bar{Q}_v}\bar{\boldsymbol{q}}_v\bar{\boldsymbol{q}}_v' \end{pmatrix}, \\ I_{3i} &= \sum_{v=0}^{D_i^s} P_{D_{i+v}^r} q_v \mathbb{E}\left[- \frac{\partial^2 \log f(Y; \eta_v, \sigma)}{\partial(\boldsymbol{\gamma}', \sigma)' \partial(\boldsymbol{\gamma}', \sigma)} \middle| \boldsymbol{x} \right], \end{aligned} \tag{2.8}$$

in which $P_u = 0$ *if* $u < 0$.

The expectations in computing the information matrix in the lemma above are taken with respect to the joint distribution of the completely observable random variables $(r, \boldsymbol{x})$. Hence, they can readily be estimated with their empirical versions. This remark also applies to all asymptotic variances appearing in latter theorems.

2.3. Parameter estimation

For every individual (i, k), use $\boldsymbol{E}_{ik}^o$ for $\boldsymbol{E}^o$ in (2.4) so that its log-likelihood $l_{E_{ik}^o}$ and score functions $\frac{\partial l_{E_{ik}^o}}{\partial \boldsymbol{\theta}}$ can be deduced by means of (2.6) and(2.7), with the quantities there replaced by their (i, k)-instances, for example, $\boldsymbol{x}$, N^s, N_u^{rbns}, p_u,

q_v, and so on by $\boldsymbol{x}_{ik}$, N_{ik}^s, N_{iku}^{rbns}, p_{iku}, q_{ikv}, and so on, respectively. Thanks to the statistical independence among policies, the overall score functions is $\frac{\partial l}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n}\sum_{k=1}^{m_i} \frac{\partial l_{\boldsymbol{E}_{ik}^o}}{\partial \boldsymbol{\theta}}$. The MLEs of the unknown parameters are thus established by equating the overall score functions to zero and can be solved by standard computing packages (e.g., we used the R package Rsolnp by Ghalanos and Theussl, 2015) in the real data analysis below). The asymptotic distributions are provided in the following theorem, in which the regularity conditions are standard and can be found in any standard textbook in statistical asymptotics, for example, van der Vaart (2000).

Theorem 2.1 (Asymptotics of MLE). *Under the usual regularity conditions,* $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}})$, $\hat{\boldsymbol{\rho}}$, *and* $(\hat{\boldsymbol{\gamma}}, \hat{\sigma})$ *are asymptotically mutually independent with:*

$$\sqrt{m}\text{vec}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) \xrightarrow{L} N\left(0, I_1^{-1}\right), \sqrt{m}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{L} N\left(0, I_2^{-1}\right) \text{ and}$$

$$\sqrt{m}\text{vec}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}, \hat{\sigma} - \sigma) \xrightarrow{L} N\left(0, I_3^{-1}\right),$$

where

$$I_1 = \mathbb{E}\left[\sum_{u=0}^{D^r}\sum_{i=1}^{n-u} \kappa_i p_u \begin{pmatrix} 1 \\ \boldsymbol{\delta}_u - \boldsymbol{p} \end{pmatrix} (1, \boldsymbol{\delta}_u' - \boldsymbol{p}') \otimes (r \exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{x}\boldsymbol{x}')\right],$$

$$I_2 = \mathbb{E}\left[\left(\sum_{i=1}^{n-D^s} \kappa_i P_{n-i-D^s}(\text{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}') \otimes (r \exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{x}\boldsymbol{x}')\right)\right]$$

$$+ \mathbb{E}\left[\left(\sum_{v=0}^{D^s-1}\sum_{i=(n-v-D^r)\vee 1}^{n-v} \kappa_i p_{n-i-v} \begin{pmatrix} \text{diag}(\boldsymbol{q}_v) - \boldsymbol{q}_v\boldsymbol{q}_v' & -\boldsymbol{q}_v\overline{\boldsymbol{q}}_v' \\ -\overline{\boldsymbol{q}}_v\boldsymbol{q}_v' & \frac{Q_v}{\overline{Q}_v}\overline{\boldsymbol{q}}_v\overline{\boldsymbol{q}}_v' \end{pmatrix}\right)\right.$$

$$\left.\otimes (r \exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{x}\boldsymbol{x}')\right] \text{ and}$$

$$I_3 = \mathbb{E}\left[r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \sum_{v=0}^{D^s}\sum_{i=1}^{n-v} \kappa_i P_{D_{i+v}^r} q_v \mathbb{E}\left[-\left.\frac{\partial^2 \log f(Y; \eta_v, \sigma)}{\partial(\boldsymbol{\gamma}', \sigma)'\partial(\boldsymbol{\gamma}', \sigma)}\right| \boldsymbol{x}\right]\right].$$

3. Loss reserving

This section precisely specifies the terminologies "loss reserve" and "loss reserving" and establishes a computation procedure for loss reserves. It then discusses the properties of loss reserving with and without individual information and thus shows the impact of introducing individual information.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

3.1. Loss reserving with individual information

From now on, the number m of the whole portfolio exposed to n accident years is explicitly indicated for the purpose of asymptotic analysis. Moreover, use $\mathscr{F}$ to stand for the updated information on the observations of the individual covariates, exposures, the R-delays of reported claims, S-delays and final payments of settled claims, and censored S-delays of RBNS claims, in n accident years.

The outstanding liabilities of the insureds—that is, the sum of future payments for all RBNS and IBNR claims—can be represented as:

$$R = R^{rbns} + R^{ibnr} = \sum_{i=1}^{n} R_i^{rbns} + \sum_{i=1}^{n} R_i^{ibnr} = \sum_{i=1}^{n} R_i^{rbns} + \sum_{i=n-D^r+1}^{n} R_i^{ibnr}, \tag{3.1}$$

where

$$R_i^{rbns} = \sum_{k=1}^{m_i} \sum_{l=1}^{N_{ik}} Y_{ikl} I_{\mathscr{A}_i^{rbns}}(U_{ikl}, V_{ikl}),\ 1 \le i \le n, \text{ and}$$
$$R_i^{ibnr} = \begin{cases} 0, & 1 \le i \le n - D^r, \\ \sum_{k=1}^{m_i} \sum_{l=1}^{N_{ik}} Y_{ikl} I_{\mathscr{A}_i^{ibnr}}(U_{ikl}, V_{ikl}), & n - D^r + 1 \le i \le n. \end{cases} \tag{3.2}$$

are, respectively, the outstanding liabilities of RBNS and IBNR claims occurring in accident year i. Then, as in Huang *et al.* (2015b), "loss reserve" is defined as the projection:

$$R_m = R_m(\boldsymbol{\theta}) = \mathbb{E}[R|\mathscr{F}] \tag{3.3}$$

of R on the information $\mathscr{F}$ that is observed at the evaluation date n, where the unknown parameter $\boldsymbol{\theta}$ is explicitly indicated as a reminder of the uncertainty in the distribution of the random element $\boldsymbol{E}$. Recall the definition of μ_vs in Assumption 1 and denote by:

$$\tilde{\mu}_v = \mathbb{E}[Y|V \ge v, \boldsymbol{x}] = \frac{\sum_{t=v}^{D^s} q_t \mu_t}{\sum_{t=v}^{D^s} q_t},\ v = 0, 1, \ldots, D^s, \tag{3.4}$$

such that one has the policy-specified quantities:

$$\mu_{ikv} = \mu_v(g^{-1}(\boldsymbol{x}'_{ikv}\boldsymbol{\gamma}), \sigma) \text{ and } \tilde{\mu}_{ikv} = \frac{\sum_{t=v}^{D^s} q_{ikt}\mu_{ikt}}{\sum_{t=v}^{D^s} q_{ikt}},\ v = 0, 1, \ldots, D^s, \tag{3.5}$$

where the function $\mu_v(\cdot)$ is defined in Assumption 1.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

The following theorem provides a formula to compute the loss reserve $R_m(\boldsymbol{\theta})$, in which O_v^{rbns} and O_u^{ibnr} respectively correspond to the reserves for the RBNS claims reported in year $n-v+1$ and the IBNR claims with an R-delay u.

Theorem 3.1. *Under Assumption 1,*

$$R_m(\boldsymbol{\theta}) = R_m^{rbns}(\boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\sigma}) + R_m^{ibnr}(\boldsymbol{\theta}) = \sum_{v=1}^{D^s} O_v^{rbns}(\boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\sigma}) + \sum_{u=1}^{D^r} O_u^{ibnr}(\boldsymbol{\theta}), \tag{3.6}$$

where $O_v^{rbns} = \sum_{i=(n-v-D^r+1)\vee 1}^{n-v+1} O_{iv}^{rbns}$ *and* $O_u^{ibnr} = \sum_{i=n-u+1}^{n} O_{iu}^{ibnr}$ *with*

- $O_{iv}^{rbns} = \sum_{k=1}^{m_i} N_{ik,n-i-v+1}^{rbns} \tilde{\mu}_{ikv}$, *the reserves for the RBNS liabilities of claims occurring in year i and reported in year* $n-v+1$ *and*
- $O_{iu}^{ibnr} = \sum_{k=1}^{m_i} p_{iku} r_{ik} \exp(\boldsymbol{x}_{ik}'\boldsymbol{\beta}) \tilde{\mu}_{ik0}$, *the reserves for the IBNR claims occurring in year i with R-delay u.*

Equation (3.6) is almost the same as Equation (3.2) in Huang *et al.* (2015b) except for the computation of the summands O_{iv}^{rbns} and O_{iu}^{ibnr} due to introducing covariates $\boldsymbol{x}_{ik}$s. Because the assumption $n = D^r + D^s + 1$ in Huang *et al.* (2015b) is replaced by $\max(D^r, D^s) < n \le D^r + D^s + 1$, however, a different computation procedure for the loss reserve $R_m(\boldsymbol{\theta})$ is required. Depicted in Table 1 below is a new procedure, for easy reference, which intuitively displays the generating processes of IBNR reserve on the right panel and RBNS reserve on the left panel. Note that claims occurring at years from 1 to $n - D^r$ have been reported at the evaluation date. We write $\underline{b}^r = n - D^r + 1$ to represent the beginning year from which R-delays (i.e., IBNR claims) need to be taken care of. The computation of IBNR reserves turns out to still have some similarity to the classical chain ladder method: there the terms in $\{O_{iu}^{ibnr} : i + u > n\}$ correspond to the blank entries in the down-right part of a run-off table. The significant difference from the chain-ladder method exists in how the entries O_{iu}^{ibnr}s are computed. Unlike IBNR claims, there are more years for which RBNS need to be taken care of. The entries in the left-lower part of the table correspond to RBNS reserves $\{O_{iv}^{rbns}; n - D^r < i + v < n\}$. Particularly, every column corresponds to an S-delay v and is made of a series of reserves for RBNS claims reported in year $n - v + 1$, that is, $\{O_{iv}^{rbns}; i = (n - v - D^r + 1) \vee 1, \ldots, n - v + 1\}$. Especially, this implies that the beginning year for S-delays (i.e., RBNS claims) to be taken care of is $\underline{b}^s = (n - D^r - D^s + 1) \vee 1 = 1$ if $n \le D^r - D^s$ and 2 if $n = D^r + D^s + 1$. Moreover, while the expression of the RBNS reserves were designed in the form of the theorem so as to give the tabular algorithm, it allows for other alternatives. For example, they can be organized in the years the claims have been reported so as to get another computation algorithm similar to the current IBNR reserve. With this alternative, one works with two run-off-like triangles.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 1

COMPUTATION OF LOSS RESERVE $R_m(\boldsymbol{\theta}) = R_m^{rbns}(\boldsymbol{\rho}, \boldsymbol{\gamma}, \sigma) + R_m^{ibnr}(\boldsymbol{\theta})$.

RBNS reserve	Settlement delay D^s	$\cdots$	$\cdots$	2	1	Accident year	Reporting delay 1	2	$\cdots$	$\cdots$	D^r	IBNR reserve
$R_{\underline{b}^s}^{rbns}$	$O_{\underline{b}^s, D^s}^{rbns}$					$\underline{b}^s$						
$\downarrow$	$\vdots$	$\ddots$				$\vdots$						
$\downarrow$	$\vdots$	$\vdots$	$\ddots$			$\vdots$						
$R_{\underline{b}^r-1-1}^{rbns}$	$O_{\underline{b}^r-2, D^s}^{rbns}$	$\leftarrow$	$\leftarrow$	$O_{\underline{b}^r-2,2}^{rbns}$		$\underline{b}^r - 2$						
$R_{\underline{b}^r-1}^{rbns}$	$O_{\underline{b}^r-1, D^s}^{rbns}$	$\leftarrow$	$\leftarrow$	$O_{\underline{b}^r-1,2}^{rbns}$	$O_{\underline{b}^r-1,1}^{rbns}$	$\underline{b}^r - 1$						
$R_{\underline{b}^r}^{rbns}$	$O_{\underline{b}^r, D^s}^{rbns}$	$\leftarrow$	$\leftarrow$	$O_{\underline{b}^r,2}^{rbns}$	$O_{\underline{b}^r,1}^{rbns}$	$\underline{b}^r$					$O_{\underline{b}^r, D^r}^{ibnr}$	$R_{\underline{b}^r}^{ibnr}$
$\downarrow$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$				$\iddots$	$\vdots$	$\downarrow$
$\downarrow$	O_{n-D^s+1, D^s}^{rbns}	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$			$\iddots$	$\vdots$	$\vdots$	$\downarrow$
R_{n-1}^{rbns}				$O_{n-1,2}^{rbns}$	$O_{n-1,1}^{rbns}$	$n - 1$		$O_{n-1,2}^{ibnr}$	$\rightarrow$	$\rightarrow$	O_{n-1, D^r}^{ibnr}	R_{n-1}^{ibnr}
R_n^{rbns}					O_{n1}^{rbns}	n	$O_{n,1}^{ibnr}$	$O_{n,2}^{ibnr}$	$\rightarrow$	$\rightarrow$	O_{n, D^r}^{ibnr}	R_n^{ibnr}
R^{rbns}	Total: $R_m = R^{rbns} + R^{ibnr}$											R^{ibnr}

Note. Arrows $\rightarrow$ and $\leftarrow$ have two purposes: indicating the direction of the summation and acting as an ellipsis sign.

Because of the uncertainty in the distribution of the random element $\boldsymbol{E}$, as what was clearly indicated in Theorem 3.1, the loss reserve $R_m(\boldsymbol{\theta})$ is still a random function of the unknown parameters $\boldsymbol{\theta}$ and thus needs to be further estimated. Accordingly, the term "loss/claims reserving" refers to two purposes: the reasonable estimate of the loss reserve and the procedure to produce that estimate. Formally, after obtaining reasonable estimates $\hat{\boldsymbol{\theta}}$, as, for example, what was done in Section 2.3, the term "loss reserving" is clearly specified below.

Definition 3.1. *The term loss reserving refers to the (random) quantity:*

$$\hat{R}_m = R_m(\hat{\boldsymbol{\theta}}) = \sum_{v=1}^{D^s} O_v^{rbns}(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\gamma}}, \hat{\sigma}) + \sum_{u=1}^{D^r} O_u^{ibnr}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\gamma}}, \hat{\sigma}), \tag{3.7}$$

obtained by substituting the estimates $\hat{\boldsymbol{\theta}}$ into $R_m(\boldsymbol{\theta})$ in (3.6) for the unknown parameters.

Therefore, $\hat{R}_m$ can also be computed by the procedure in Table 1 with the unknown parameters replaced by their estimates.

3.2. Loss reserving neglecting individual information

For every individual (i, k), when the information on $x_{ik1}, \ldots, x_{ik,d-1}$ is neglected, the loss reserve can be obtained by simply replacing with zero the coefficients of $x_{ikj}, j = 1, 2, \ldots, d-1$ in Theorem 3.1. The resulting reserve resembles but still slightly different from what was obtained by Huang *et al.* (2016), due to the parametric form of the final payments' distribution used here, as precisely described below.

On the one hand, at the population level, ignoring the individual information $\boldsymbol{x}$ produces a false model (parallel to Distribution Assumption 1) that holds true only when the coefficients of $x_{ikj}, j = 1, 2, \ldots, d-1$ are exactly zero.

Distribution Assumption 2 (The false model of $\boldsymbol{E}$).

(1) The expectation of claim numbers per exposure becomes $\lambda = \exp(\beta_0)$.
(2) The probabilities of R-delay U and S-delay V for each claim become

$$p_u = \frac{\exp(\pi_{u0})}{\sum_{j=0}^{D^r} \exp(\pi_{j0})}, u = 0, 1, \ldots, D^r \text{ and } q_v = \frac{\exp(\rho_{v0})}{\sum_{j=0}^{D^s} \exp(\rho_{j0})}, v = 0, 1, \ldots, D^s. \tag{3.8}$$

(3) Given the S-delay $V = v$, the final payment Y is distributed as a density function $f(y; \gamma_v^h, \sigma)$, where γ_v^h corresponds to $\gamma_0 + \gamma_{d-1+v}$ for $v \geq 1$ and γ_0 for $v = 0$, writing $\boldsymbol{\gamma}^h = (\gamma_0^h, \gamma_1^h, \ldots, \gamma_{D^s}^h)$.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

The loss reserve/reserving is then derived from this false model with its "unknown parameters" λ, $(p_0, p_1, \ldots, p_{D^r})$, $(q_0, q_1, \ldots, q_{D^s})$, and $(\boldsymbol{\gamma}^h, \sigma)$. It turns out helpful to use alternatively the expected claim numbers per exposure with R-delay u:

$$\lambda_u = \lambda p_u, u = 0, 1, \ldots, D^r, \text{that is, } \lambda = \sum_{u=0}^{D^r} \lambda_u \text{ and } p_u = \lambda_u/\lambda, u = 0, 1, \ldots, D^r. \tag{3.9}$$

On the other hand, at the sample level, the intensity $r_{ik} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\beta})$ of claim occurrence per exposure, R- and S-delays, final payments, and so on are taken to be homogeneous over policies but with incorrectly assigned distributions, because once again the coefficients of $x_{ik1}, x_{ik2}, \ldots x_{ik,d-1}$ are incorrectly assigned zeros. Especially, $\tilde{\mu}_{ikv}$ and $p_{iku} \exp(\boldsymbol{x}'_{ik}\boldsymbol{\beta})$ in (3.6) reduce to $\tilde{\mu}_v$ and λ_u, respectively, such that the outstanding liabilities are now simplified to:

$$O_v^{rbns} = \sum_{i=(n-v-D^r+1)\vee 1}^{n-v+1} \sum_{k=1}^{m_i} N_{ik,n-i-v+1}^{rbns} \tilde{\mu}_v := G_v \tilde{\mu}_v \text{ and}$$
$$O_u^{ibnr} = \sum_{i=n-u+1}^{n} \sum_{k=1}^{m_i} r_{ik} \lambda_u \tilde{\mu}_0 := r_{[u]} \lambda_u \tilde{\mu}_0,$$

where

$$G_v = \sum_{i=(n-v-D^r+1)\vee 1}^{n-v+1} \sum_{k=1}^{m_i} N_{ik,n-i-v+1}^{rbns}, v = 1, 2, \ldots, D^s \text{ and}$$
$$r_{[u]} = \sum_{i=n-u+1}^{n} \sum_{k=1}^{m_i} r_{ik}, u = 1, 2, \ldots, D^r. \tag{3.10}$$

Note that G_v is the total number of RBNS claims reported in year $n - v + 1$. Also introduce

$$r_{(u)} = \sum_{i=1}^{n-u} \sum_{k=1}^{m_i} r_{ik}, u = 0, 1, \ldots, D^r. \tag{3.11}$$

Remark 3.1. *The model in Assumption 2 is called false because it is not the true distribution of the loss development. Denote by $\mathbb{E}_{\boldsymbol{x}}$ the expectation operation with respect to the distribution of $\boldsymbol{x}$. Specifically, if individual information is neglected from the analysis, then, given the S-delay $V = v$, the true density of the payment Y is $\mathbb{E}_{\boldsymbol{x}}[f(y; \eta_v, \sigma)]$, rather than the incorrectly taken $f(y; \gamma_v^h, \sigma)$ in Assumption 2. The terms "false loss reserve" and "false loss reserving" are also used below, with similar meanings.*

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

The above procedure simply implies the following theorem that is straightforwardly deduced from Theorem 3.1. The subscript "H" indicates "Huang *et al.*" hereafter.

Theorem 3.2. *The false loss reserve obtained by neglecting individual information* $\{\boldsymbol{x}_{ik}\}$ *is* $R_H = R_m^{rbns} + R_m^{ibnr} = \sum_{v=1}^{D^s} G_v\tilde{\mu}_v + \tilde{\mu}_0 \sum_{u=1}^{D^r} r_{[u]}\lambda_u$.

In order to estimate the "unknown parameters" $\tilde{\mu}_v$ and λ_u in Theorem 3.2, the false log-likelihood is then deduced from the true one by again omitting the individual information. To this end, write

$$\tilde{N}_u^r = \sum_{i=1}^{n-u}\sum_{k=1}^{m_i} N_{iku}^r,\ u=0,1,\ldots,D^r,\ \tilde{N}_v^s = \sum_{i=1}^{n-v}\sum_{k=1}^{m_i} N_{ikv}^s,\ v=0,1,\ldots,D^s, \quad (3.12)$$

to represent the total number of reported claims with R-delay u and the total number of settled claims with S-delay v, respectively. Then, set the coefficients of $x_{ikj}, j=1,2,\ldots,d-1$ in (2.6) to zero and, as usually done in survival analysis, reparameterize $\{q_v\}$ as the harzard rates:

$$h_v = \Pr(V=v|V\geq v) = \frac{q_v}{\bar{Q}_{v-1}},\ \ v=0,1,\ldots,D^s, \text{that is, } q_0=h_0 \text{ and}$$
$$q_v = h_v\prod_{s=0}^{v-1}(1-h_s),\ v=1,\ldots,D^s. \quad (3.13)$$

As a result, the false log-likelihood of the "unknown parameters" can be represented as:

$$l_H = \sum_{u=0}^{D^r}\left(\tilde{N}_u^r\log\lambda_u - r_{(u)}\lambda_u\right) + \sum_{v=0}^{D^s-1}\left(\tilde{N}_v^s\log h_v + (\tilde{\mathcal{N}}_v^r - \tilde{N}_v^s)\log(1-h_v)\right)$$
$$+\sum_{v=0}^{D^s}\sum_{l=1}^{\tilde{N}_v^s}\log f(Y_{vl};\gamma_v^h,\sigma),$$

where, with N_{ikuv} the numbers of policy (i,k)'s claims with R- and S-delay (u,v):

$$\tilde{\mathcal{N}}_v^r = \sum_{i=1}^{n-v}\sum_{u=0}^{D_{i+v}^r}\sum_{s=v}^{D^s}\sum_{k=1}^{m_i} N_{ikuv} \quad (3.14)$$

is the total number of reported claims with S-delay no less than v, satisfying $\tilde{\mathcal{N}}_v^r = \sum_{t=v}^{D^s}\tilde{N}_t^s + \sum_{t=v+1}^{D^s} G_t$. The "MLE"s are therefore,

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

$$\begin{cases} \hat{\lambda}_u = \dfrac{\tilde{N}_u^r}{r_{(u)}}, u = 0, 1, \ldots, D^r, \\ \hat{h}_v = \dfrac{\tilde{N}_v^s}{\tilde{\mathcal{N}}_v^r}, v = 0, 1, \ldots, D^s \text{ (so that, } \hat{q}_0 = \hat{h}_0 \text{ and } \hat{q}_v = \hat{h}_v \prod_{s=0}^{v-1} (1 - \hat{h}_s), v = 1, \ldots, D^s \text{ by (3.13))}, \\ (\hat{\boldsymbol{\gamma}}^h, \hat{\sigma}) = \text{the solution of} \sum_{v=0}^{D^s} \sum_{l=1}^{\tilde{N}_v^s} \dfrac{\partial \log f(Y_{vl}; \gamma_v^h, \sigma)}{\partial(\boldsymbol{\gamma}^{h\prime}, \sigma)'} = 0 \text{ (so that, } \hat{\mu}_v = \mu_v(g^{-1}(\hat{\gamma}_v^h), \hat{\sigma}), v = 0, 1, \ldots, D^s) \end{cases} \tag{3.15}$$

With equality (3.4), estimate $\tilde{\mu}_v$ by $\hat{\tilde{\mu}}_v = \frac{\sum_{s=v}^{D^s} \hat{q}_s \hat{\mu}_s}{\sum_{s=v}^{D^s} \hat{q}_s}$. The following theorem is simply obtained by replacing the "unknown parameters" $\tilde{\mu}_v$ and λ_u in Theorem 3.2 with their estimates $\hat{\tilde{\mu}}_v$ and $\hat{\lambda}_u$.

Theorem 3.3. *The false loss reserving obtained by neglecting the individual information is* $\hat{R}_H = \sum_{v=1}^{D^s} G_v \hat{\tilde{\mu}}_v + \hat{\tilde{\mu}}_0 \sum_{u=1}^{D^r} r_{[u]} \hat{\lambda}_u$.

3.3. Asymptotic behaviors of the loss reservings

The distributional features of the deviations $\hat{R}_m - R_m$ and $\hat{R}_H - R_m$ demonstrate the effects of introducing covariates $\boldsymbol{x}$, as shown separately in the following two theorems in asymptotic sense.

Theorem 3.4 needs the notation: $\overline{\boldsymbol{\mu}}_v = (\mu_{v+1}, \mu_{v+2}, \ldots, \mu_{D^s})'$ and $\boldsymbol{\alpha}_v = (q_v \frac{\partial \mu_v}{\partial(\boldsymbol{\gamma}', \sigma)}, q_{v+1} \frac{\partial \mu_{v+1}}{\partial(\boldsymbol{\gamma}', \sigma)}, \ldots, q_{D^s} \frac{\partial \mu_{D^s}}{\partial(\boldsymbol{\gamma}', \sigma)})'$, $v = 0, \ldots, D^s$, where $\frac{\partial \mu_v}{\partial(\boldsymbol{\gamma}', \sigma)} = (\frac{\partial \mu_v}{\partial \eta_v} \frac{\boldsymbol{x}_v'}{\dot{g}(\eta_v)}, \frac{\partial \mu_v}{\partial \sigma})$ with $\dot{g}$ the derivative function of g.

Theorem 3.4. $\frac{1}{\sqrt{m}}(\hat{R}_m - R_m) \xrightarrow{L} N(0, \sigma_I^2)$ *with* $\sigma_I^2 = \sum_{j=1}^3 \tilde{g}_j' I_j^{-1} \tilde{g}_j$, *where* I_j*s are defined as in Theorem 2.1 and*

$$\tilde{g}_1 = \sum_{u=0}^{D^r} \sum_{i=n-u+1}^{n} \kappa_i \mathbb{E} \left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \tilde{\mu}_0 \begin{pmatrix} 1 \\ \boldsymbol{\delta}_u - \boldsymbol{p} \end{pmatrix} \otimes \boldsymbol{x} \right],$$

$$\tilde{g}_2 = \sum_{u=0}^{D^r} \sum_{i=(n-u-D^s+1)\vee 1}^{n} \kappa_i \mathbb{E} \left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \begin{pmatrix} \mathbf{0} \\ \mathrm{diag}(\overline{\boldsymbol{q}}_{(n-i-u)_+}) \overline{\boldsymbol{\mu}}_{(n-i-u)_+} - \tilde{\mu}_{(n-i-u+1)_+} \overline{\boldsymbol{q}}_{(n-i-u)_+} \end{pmatrix} \otimes \boldsymbol{x} \right],$$

$$\tilde{g}_3 = \sum_{u=0}^{D^r} \sum_{i=(n-u-D^s+1)\vee 1}^{n} \kappa_i \mathbb{E} \left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_{(n-i-u+1)_+} \end{pmatrix} \right]. \tag{3.16}$$

The misspecification of the model giving $\hat{R}_H$ brings high complexity to the notation and makes it difficult to express and prove the asymptotic results

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

commensurate with Theorem 3.4. First note that $\hat{R}_H$ is a function of the statistics $r_{(u)}$s in (3.11), $\tilde{N}_u^r$s and $\tilde{N}_v^s$s in (3.12), $\tilde{\mathcal{N}}_v^r$s in (3.14), and $(\hat{\boldsymbol{\gamma}}^h, \hat{\sigma})$ through $\hat{\mu}_v = \mu_v(g^{-1}(\hat{\gamma}_v^h), \hat{\sigma})$. The law of large numbers readily gives

$$\begin{cases} \frac{r_{[u]}}{m} \stackrel{a.s.}{\to} \check{r}_{[u]} := \sum_{i=n-u+1}^{n} \kappa_i \mathbb{E}[r], u = 1, 2, \ldots, D^r, \\ \hat{\lambda}_u \stackrel{a.s.}{\to} \check{\lambda}_u := \dfrac{\mathbb{E}[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta})]}{\mathbb{E}[r]}, u = 1, 2, \ldots, D^r, \\ \hat{h}_v \stackrel{a.s.}{\to} \check{h}_v := \dfrac{\sum_{i=1}^{n-v} \kappa_i \mathbb{E}\left[P_{D_{i+v}^r} r \exp(\boldsymbol{x}'\boldsymbol{\beta}) q_v\right]}{\sum_{i=1}^{n-v} \kappa_i \mathbb{E}\left[P_{D_{i+v}^r} r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \bar{Q}_{v-1}\right]}, v = 0, 1, \ldots, D^s, \end{cases} \tag{3.17}$$

where r is the exposure of a representative policy, as postulated in Assumption 1. Let

$$\left(\check{\boldsymbol{\gamma}}^{h\prime}, \check{\sigma}\right) = \underset{(\boldsymbol{\gamma}^h, \sigma)}{\text{Argmax}} \sum_{v=0}^{D^s} \sum_{i=1}^{n-v} \kappa_i \mathbb{E}\left[r \exp(\boldsymbol{x}'\boldsymbol{\beta}) P_{D_{i+v}^r} q_v \log f\left(Y; \gamma_v^h, \sigma\right)\right]. \tag{3.18}$$

Denote by

$$\check{\mu}_v = \mu_v\left(g^{-1}(\check{\gamma}_v^h), \check{\sigma}\right) \text{ and } \check{q}_v = \check{h}_v \prod_{s=0}^{v-1}\left(1 - \check{h}_s\right) \tag{3.19}$$

the limits of $\hat{\mu}_v$ and $\hat{q}_v$ as $m \to \infty$. Then, corresponding to $\hat{R}_H$, we introduce

$$\check{R}_H = \sum_{v=1}^{D^s} G_v \check{\tilde{\mu}}_v + \check{\tilde{\mu}}_0 \sum_{u=1}^{D^r} r_{[u]} \check{\lambda}_u, \text{ where } \check{\tilde{\mu}}_v = \sum_{s=v}^{D^s} \check{q}_s \check{\mu}_s \Big/ \sum_{s=v}^{D^s} \check{q}_s \tag{3.20}$$

and define

$$\sigma_H^2 = \xi' \check{\Sigma} \xi \text{ with } \xi = \left(\check{\tilde{\mu}}_0(\check{r}_{[1]}, \check{r}_{[2]}, \ldots, \check{r}_{[D^r]}), \sum_{u=1}^{D^r} \check{r}_{[u]} \check{\lambda}_u, \check{G}_1, \check{G}_2, \ldots, \check{G}_{D^s}\right)', \tag{3.21}$$

where $\check{G}_v = \sum_{u=0}^{D^r} \kappa_{n-u-v+1} \mathbb{E}\left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \bar{Q}_{v-1}\right]$, $v = 0, 1, \ldots, D^s$, and the meaning of $\check{\Sigma}$ is deferred to (A14). With this notation, the following theorem presents $\hat{R}_H$'s asymptotic behaviors.

Theorem 3.5. *Under usual regular conditions on $f(y; \eta, \sigma)$ (see White, 1982), with parameters $\check{R}_H$ in (3.20) and σ_H^2 in (3.21):*

$$\frac{1}{\sqrt{m}}\left(\hat{R}_H - \check{R}_H\right) \stackrel{L}{\to} N\left(0, \sigma_H^2\right) \text{ and } \frac{1}{m}\left(\check{R}_H - R_m\right) \to \Delta, \tag{3.22}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

where the asymptotic bias is

$$\Delta = \sum_{v=1}^{D^s} \sum_{u=0}^{D^r_v} \kappa_{n-v-u+1} \mathbb{E}\left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) \bar{Q}_{v-1}(\check{\tilde{\mu}}_v - \tilde{\mu}_v) \right]$$

$$+ \sum_{u=1}^{D^r} \sum_{i=n-u+1}^{n} \kappa_i \mathbb{E}\left[p_u r \exp(\boldsymbol{x}'\boldsymbol{\beta}) (\check{\tilde{\mu}}_0 - \tilde{\mu}_0) \right]. \tag{3.23}$$

This section concludes with the following two obvious but important facts deduced from Theorem 3.5:

(1) If none of the covariates has effects on the occurrence and claims development, then we have that $\check{\lambda}_u = \lambda_u = p_u \exp(\beta_0)$ (by (3.17)), $\check{q}_v = q_v$ and $\check{\mu}_v = \mu_v$ (by (3.19)), where p_u, q_v, and μ_v were defined in by (3.8), $u = 0, 1, \ldots, D^r$, $v = 0, 1, \ldots, D^s$. This results in an equality $\check{R}_H = R_m$ and thus indicates that the loss reserving $\hat{R}_H$ generally gives more efficient evaluation (e.g., also asymptotically unbiased but with smaller variance) of the loss reserves than $\hat{R}_m$.

(2) Generally, omitting the useful covariates in the loss reserving results in an asymptotic bias:

$$\frac{1}{m}\left(\hat{R}_H - R_m\right) = \frac{1}{\sqrt{m}} \frac{1}{\sqrt{m}} \left(\hat{R}_H - \check{R}_H\right) + \frac{1}{m}\left(\check{R}_H - R_m\right) \xrightarrow{p} \Delta, \tag{3.24}$$

that is, the difference $(\hat{R}_H - R_m) = O_p(m\Delta)$ is of the same order as the exposure size m, much higher than the order $\sqrt{m}$ of $\hat{R}_m - R_m$. The effects of using individual information are quite shocking compared with those demonstrated in Huang *et al.* (2016), in which the improvement of the individual loss data method over the classical chain-ladder method was only in the order of $\sqrt{m}$. Nevertheless, it is not easy to learn how Δ changes over the parameters. Some insight might be revealed by the following numerical computation conducted in the simplest setting $n = 3$, $D^r = 1$, and $D^s = 1$ years with two examples of parameter varying:

Example 1. Dimension $d = 3$ and the parameters varied in an auxiliary parameter t ranging in $[-1, 1]$ by step 0.01 as:

$$\boldsymbol{\beta}' = (2, 0.6t, -0.7t), \boldsymbol{\pi}' = (1, -0.8t, 0.8t), \boldsymbol{\rho}' = (1, -0.5t, 0.4t)', \boldsymbol{\gamma}' = (2, 0.9t, -0.6t, 0.7), \sigma = 4. \tag{3.25}$$

Example 2. Dimension $d = 4$ and parameters varied over t ranging in $[-2, 2]$ by step 0.02 as:

$$\boldsymbol{\beta}' = (1, -0.5t, 0.3t, 0.4t), \boldsymbol{\pi}' = (-1, -0.3t, 0.5t, 0.4t), \boldsymbol{\rho}' = (1, 0.4t, -0.6t, 0.8t)', \boldsymbol{\gamma}' = (3, 0.7t, -0.2t, -0.6t, 0.9), \sigma = 1. \tag{3.26}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 2

PARAMETER SETTINGS FOR THE SIMULATION.

	Settings			
Parameters	I	II	III	IV
$\boldsymbol{\beta}'$	(2, 0.6, −0.7)	(1, −0.5, 0.3, 0.4)	(2, 0, 0)	(1, 0, 0, 0)
$\boldsymbol{\pi}'$	(1, −0.8, 0.8)	(− 1, −0.3, 0.5, 0.4)	(1, 0, 0)	(− 1, 0, 0, 0)
$\boldsymbol{\rho}'$	(1, −0.5, 0.4)	(1, 0.4, −0.6, 0.8)	(1, 0, 0)	(1, 0, 0, 0)
$\boldsymbol{\gamma}'$	(2, 0.9, −0.6, 0.7)	(3, 0.7, −0.2, −0.6, 0.9)	(2, 0, 0, 0.7)	(3, 0, 0, 0, 0.9)
σ	4	1	4	1

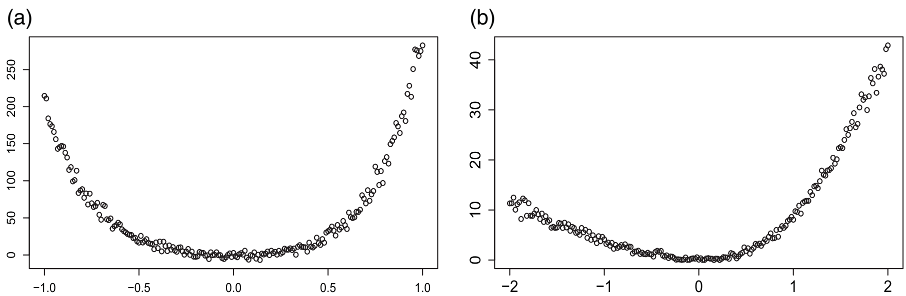

FIGURE 1: The asymptotic bias over varying coefficients of covariates. (a) Example 1. (b) Example 2.

The other details of the computation were the same as the one described in Section 3.4 below. The bias Δ was approximated using (3.24) by Monte Carlo method with $\mathbf{m} = (10,000, 10,000, 10,000)$ risk exposures, that is, 10,000 exposures every year. For every combination of the values of the parameters, with $\hat{R}_H$ computed according to Theorem 3.3 and R_m computed by inserting the parameters into (3.6), Δ was approximated by $\frac{\hat{R}_H - R_m}{m}$. In the computation, as shown by the plots of $|\Delta|$ over t in Figure 1, $|\Delta|$ increased when t got large and approached zero when the coefficients tended to zero.

3.4. A simulation study

This subsection reports the results from a small simulation study that demonstrated the impact of individual information on loss reserving with finite portfolio sizes. The simulation proceeded with $n = 3$ years, $D^r = 1$, and $D^s = 1$ under the two settings of risk exposures: $\mathbf{m}_1 = (4000, 4000, 4000)$ (4000 exposures every year) and $\mathbf{m}_2 =$ (10,000, 10,000, 10,000) (10,000 exposures every year). Four settings of the true values of the parameters as listed in Table 2 were examined: settings I ($d = 3$) and II ($d = 4$) represented the mechanism that the individual information had effects on the claim developments and III and

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 3

MEAN ± STANDARD DEVIATION OF THE RELATIVE ERRORS.

Portfolio	Case	$\hat{R}_m$	$\hat{R}_H$
$\mathbf{m}_1$	I	1.8002 ± 1.2967	40.073 ± 7.0266
	II	1.2066 ± 0.9012	18.791 ± 3.1496
	III	2.0024 ± 1.4578	1.9972 ± 1.4560
	IV	1.1663 ± 0.9030	1.1683 ± 0.8963
$\mathbf{m}_2$	I	1.0762 ± 0.7947	39.827 ± 5.0717
	II	0.7075 ± 0.4988	18.711 ± 1.9492
	III	1.2252 ± 0.9544	1.2247 ± 0.9527
	IV	0.6818 ± 0.5238	0.6812 ± 0.5246

IV represented the mechanism that the individual information had no effects by assigning 0 to the coefficients of the covariates x_{ik1}, x_{ik2} in setting I and $x_{ik1}, x_{ik2}, x_{ik3}$ in setting II. Note that $\gamma_3 = 0.7$ in settings I and III and $\gamma_4 = 0.9$ in settings II and IV characterized the correlation between the S-delays and the final claims. The risk exposures associated with every individuals were independent and identically drawn from a uniform distribution on [0, 1], the covariates were produced using standard normal distributions of dimension 2 for settings I and III and 3 for settings II and IV, and the claim amounts were drawn from a log-normal distribution whose parameters depended on $\boldsymbol{\gamma}$ and σ as defined in Distribution Assumption 1.

A total of 200 duplicates were carried out for both the portfolios $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ with parameter values specified above. In order to show the effects of individual information, for the reserving with and without $x_1, \ldots, x_{d-1}$, we computed R_m according to (3.6) with the imposed parameter values for every portfolio and then the absolute relative errors $\frac{|\hat{R}_m - R_m|}{R_m} \times 100$ and $\frac{|\hat{R}_H - R_m|}{R_m} \times 100$, for every portfolio and settings of parameters. The results from the 200 runs were summarized in the form "mean ± sd (standard deviation)" in Table 3. Comparing columns $\hat{R}_m$ and $\hat{R}_H$ clearly shows the following two facts that are coincide with the asymptotic conclusions:

(1) Under settings I and II, the individual information model (IIM) provided much more accurate loss reserving in terms of both the mean (measuring the average of the relative errors) and the standard error (measuring the stability of the relative errors).
(2) Under settings III and IV, the IIM provided loss reservings that were a bit worse in term of the standard error.

The findings can also be much more clearly depicted by the empirical densities of the 200 simulated relative errors, as shown for portfolio $\boldsymbol{m}_2$ in Figure 2. The plots illustrated that when real effects of the covariates exist (settings I and II), the absolute relative errors of $\hat{R}_H$ had both greater mean and dispersion, whereas, in the absence of real effects (settings III and IV), the two models

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

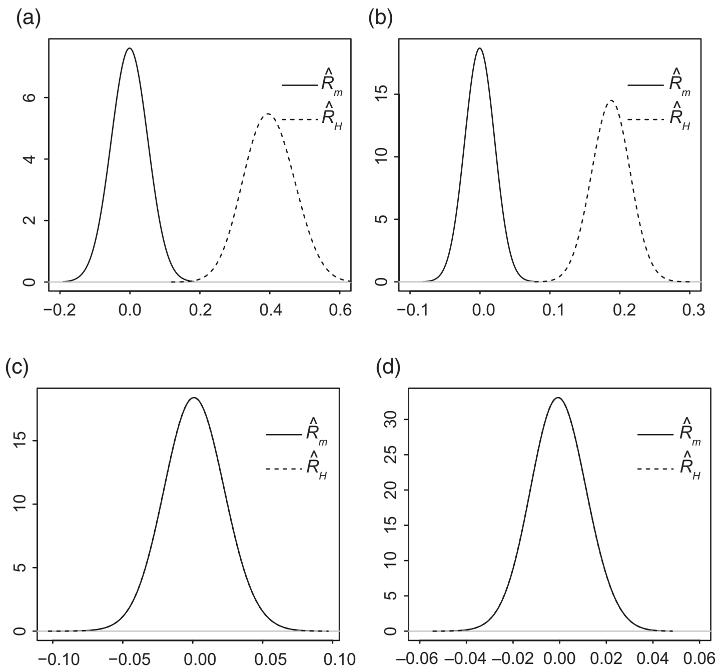

FIGURE 2: Estimated densities of the relative differences $\frac{\hat{R}_m - R_m}{R_m}$ and $\frac{\hat{R}_H - R_m}{R_m}$. (a) Setting I. (b) Setting II. (c) Setting III. (d) Setting IV.

showed almost the same prediction capability. This simulation thus supports the use of the IIM.

4. Real data application

This section addresses a real analysis in health insurance, using a dataset from a commercial insurance company in China. This section also discusses the difference between the loss reservings with and without individual information, where the former is referred to as IIM and the latter the individual data model (IDM), following the terminology in Huang *et al.* (2016).

4.1. Data description

The dataset documented the starting and ending dates of the insurance policies, and the following individual information on insureds: age, gender, policy type (serious illness insurance and general health insurance), geometric region, and occurrences and developments of claims between January 1 2019 and September 23 2019.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 4

THE INDIVIDUAL INFORMATION IN REAL DATA ANALYSIS.

Covariate	Description	Type	Levels
x_1	Insured's age	Quantitative	
x_2	Insured's gender	Binary	Female: 1, male: 0
x_3	Policy type	Binary	General health insurance: 1 Serious illness insurance: 0
$x_4 - x_8$	Geographical location	Categorical	Regions I–V: one-hot encoding with $(x_4, x_5, \ldots, x_8)$ so that region VI: $x_4 = \cdots = x_8 = 0$

The four factors were organized into eight features $x_1, \ldots, x_8$, as shown in Table 4. One month was taken as the time unit ("accident year" in previous sections). Because the highly developed communication system, the R-delays in health insurance are generally mediate; in the data, all the delays were not more than 150 days (5 months). By China Banking and Insurance Regulatory Commission, the reported claims in health insurance are generally required to be settled within 2 months if no disagreement exists. Thus, the maximum R- and S-delays were safely set to $D^r = 5$ and $D^s = 3$ (the real data supported this assumption). In order to examine the accuracy of the loss reservings, June 30 2019 was set as evaluation date. Accordingly, the dataset was partitioned into observed set (reported or settled claims and their developments before June 30 2019) and validation set (the claims and their development incurred before June 30 2019 and reported or settled after June 30 2019), that is, we worked with $n = 6$, $D^r = 5$, and $D^s = 3$ (months).

4.2. Heterogeneity of claims developments

In order to explore possible heterogeneity among risks, we produced from the data a sequence of plots first, as listed in Figure 3. Plots (a)–(c) were the daily risk exposures (effective policies/1000), the claims intensities (number of reported claims per exposure), and risk intensities (proportion of individuals over 55 years of age), all over dates, respectively. An obvious increasing in risk exposures was seen in Plot (a). The second plot clearly depicted a heterogeneity in claims intensities over time. Though there appeared a synchrony between the daily exposures and the claims intensities, it is a common sense that no dependence should exist between the two quantities. The decreasing trend from July 2019 to September 23 2019 might be attributable to IBNR claims (R-delays). The increasing trend before July 2019 might be attributable to certain other factors. Plot (c) appeared to support this by an obvious synchronous increasing between claims intensities and risk intensities and thus implied a possibility to use age of the insureds to improve projection of the outstanding claims. The last showed the violin plots of the logarithms of payments of

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

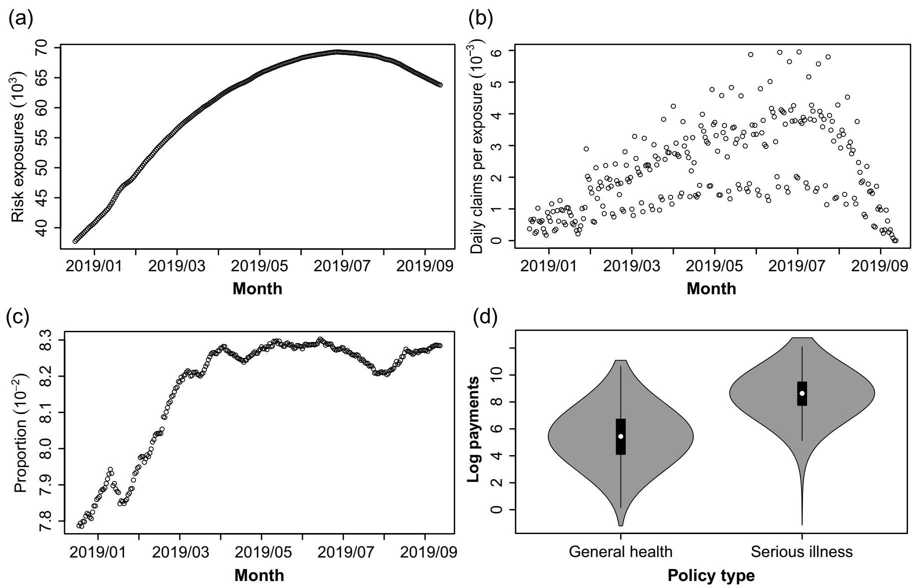

FIGURE 3: Heterogeneity of claims development. (a) Risk exposures (b) Claims intensity (c) Risk intensity (d) Logarithm of payments.

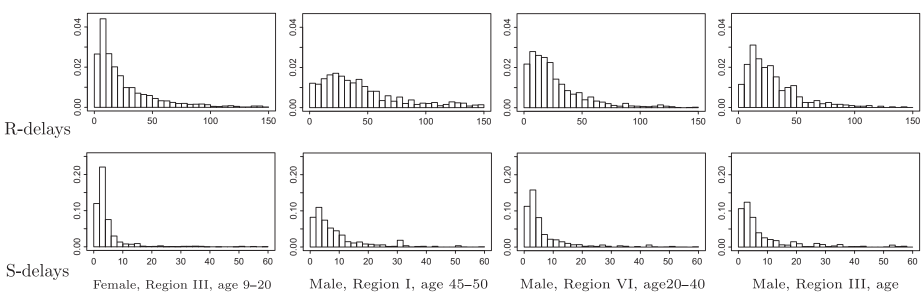

FIGURE 4: Histograms of reporting (lower) and settlement (upper) delays (in days).

the two policy types, which indicated that claims from serious illness insurance contracts generally caused larger payments.

Secondly, the histograms of R- and S-delays measured in days were also provided under a few combinations of covariate values including gender, geographical location, and age, as presented in Figure 4. It was strongly proposed that the individual information had impacts on the distributions of R- and S-delays.

Another evidence supporting claims development heterogeneity among individuals came from the model-checking procedure. This involved a sequence of statistical tests for the hypotheses that the regression coefficients were zero, after establishing the model (as illustrated in the following Subsection 4.3). In Table 5, brought forward highlighted the results from a set of such tests for

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 5

LIKELIHOOD RATIO TESTS FOR EFFECTS OF INDIVIDUAL FEATURES ON CLAIMS DEVELOPMENT.

	Hypothesis H_0	χ^2 statistic	df	p-Value
Separate	$H_{01}: \beta_k = 0, k = 1, \ldots, 8$	5166.9	8	0.0000
	$H_{02}: \pi_{uk} = 0, u = 1, \ldots, 5, k = 1, \ldots, 8$	428.9	40	0.0000
	$H_{03}: \rho_{vk} = 0, v = 1, \ldots, 3, k = 1, \ldots, 8$	161.7	24	0.0000
	$H_{04}: \gamma_k = 0, k = 1, \ldots, 8$	9439.2	8	0.0000
Joint	H_0: $H_{01} - H_{04}$ are all true	18,049.1	80	0.0000

the impacts of the covariates on the distributions of R-delays, S-delays, and the claims amounts separately (the first four tests) and jointly (the last test). The test statistics were the standard likelihood ratios (LR): $-2\log \text{LR} \sim \chi^2_k$ of degree of freedom k. From the results, one arrived at the same conclusion regarding heterogeneity.

4.3. The model and its estimation

Here addressed are the models for claims payments and parameter estimation in the established model.

When fitting claims payments given an S-delay v, three candidate models were considered:

- Pareto II: $f(y; \eta_v, \sigma) = \dfrac{\sigma \eta_v^\sigma}{(y + \eta_v)^{\sigma+1}}$, with $\log(\eta_v) = \boldsymbol{x}_v'\boldsymbol{\gamma}$ and $\mu_v = \frac{\eta_v}{\sigma-1}$, $\sigma > 1$.
- Log-normal: $f(y; \eta_v, \sigma) = \dfrac{1}{y\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(\log y - \eta_v)^2}{2\sigma^2}\right)$, with $\eta_v = \boldsymbol{x}_v'\boldsymbol{\gamma}$ and $\mu_v = \exp\left(\eta_v + \frac{\sigma^2}{2}\right)$.
- Gamma: $f(y; \eta_v, \sigma) = \dfrac{y^{\sigma-1}\exp(-y/\eta_v)}{\Gamma(\sigma)\eta_v^\sigma}$, with $\log(\eta_v) = \boldsymbol{x}_v'\boldsymbol{\gamma}$ and $\mu_v = \sigma\eta_v$,

in which the Pareto II and log-normal models are heavy-tailed, and the Gamma is regular. In order to examine the goodness of fit of those distributions, the corresponding statistics were computed, including log-likelihoods, Akaike information criterions (AICs), Bayesian information criterions (BICs), and p-values of KS, χ^2 (using R package "stats" with 20 equiprobable intervals) and AD (using R package "ADGofTest") tests for every models, see the left panel of Table 6. All the statistics supported to use Pareto II to fit the data. The same statistics were also computed for the corresponding models ignoring the individual information, and these statistics were shown in the right panel of Table 6. The results provided strong evidence against using those false models.

As an auxiliary tool, a graphical diagnostic was also conducted for the candidate models by checking the Q-Q plots of $F(Y_l; \hat{\eta}_v, \hat{\sigma})$s. Should the model be true, these statistics would approximately follow a uniform distribution on [0, 1], where F is the hypothesized cumulative distribution function of the

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 6

SELECTION STATISTICS FOR CLAIMS PAYMENTS AND p-VALUE (IN THE BRACKETS) OF GOODNESS-OF-FIT STATISTICS.

	Model selection statistics					
	With individual information			Without individual information		
Distribution	Log-likelihood (KS test)	AIC (χ^2 test)	BIC (AD test)	Log-likelihood (KS test)	AIC (χ^2 test)	BIC (AD test)
Pareto II	**−111,699.3**	**223,424.6**	**223,520.8**	−116, 187.9	232,385.7	232,422.7
	(**0**.6342)	(**0**.1037)	(**0**.1398)	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-7})$
Log-normal	−111, 856.5	223,739.0	223,835.2	−115, 658.7	231,327.4	231,364.5
	(1.5×10^{-11})	$(< 10^{-15})$	$(< 10^{-7})$	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-7})$
Gamma	−112, 807.2	225,640.3	225,640.3	−115, 419.3	230,848.6	230,885.6
	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-7})$	$(< 10^{-15})$	$(< 10^{-15})$	$(< 10^{-7})$

The boldface in Table 6 is to highlight the best largest values of model selection statistics computed for Pareto II, log-normal and gamma distribution.

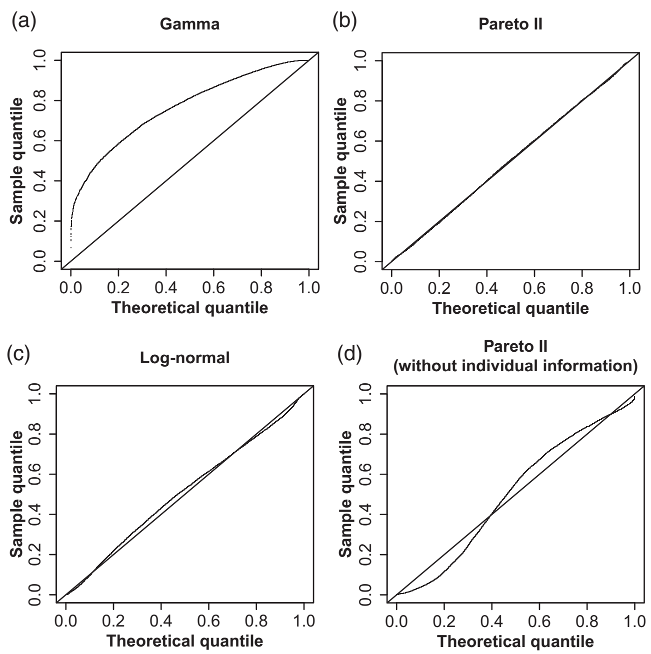

FIGURE 5: Q-Q plots for Gamma, Pareto II, and log-normal distributions. (a) Gamma (b) Pareto II (c) Log-normal (d) Parate II under IDM

claims payment given S-delay v and Y_ls are the observations of payments accompanied with S-delay v. The Q-Q curve of the IDM with Pareto II was also plotted and placed to the right of Figure 5. Clearly, the Q-Q analysis displayed in plots (a) to (c) again strongly recommended IIM with Pareto II distributed claims payments.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 7

ESTIMATED PARAMETERS, STANDARD ERRORS, AND P-VALUES FOR CLAIMS NUMBERS: IIM.

	Estimate	Std. error	p-Value	Estimate	Std. error	p-Value	Estimate	Std. error	p-Value
		β			π_1			π_2	
Intercept	−4.8773	0.0083	0.0000	1.0763	0.0166	0.0000	0.1860	0.0234	0.0000
x_1	0.0316	0.0001	0.0000	0.0067	0.0003	0.0000	0.0090	0.0005	0.0000
x_2	0.0595	0.0117	0.0000	−0.0537	0.0234	0.0216	−0.0653	0.0330	0.0479
x_3	1.2176	0.0123	0.0000	−0.1393	0.0246	0.0000	−0.0933	0.0347	0.0071
x_4	0.7673	0.0244	0.0000	−0.2000	0.0482	0.0000	0.0982	0.0592	0.0970
x_5	0.6899	0.0342	0.0000	−0.3350	0.0679	0.0000	−0.1250	0.0842	0.1378
x_6	0.4640	0.0130	0.0000	−0.8323	0.0263	0.0000	−1.0009	0.0387	0.0000
x_7	0.3713	0.0150	0.0000	−0.6942	0.0301	0.0000	−0.9388	0.0448	0.0000
x_8	0.2327	0.0279	0.0000	−0.6831	0.0560	0.0000	−0.7452	0.0773	0.0000
		π_3			π_4			π_5	
Intercept	−0.7168	0.0400	0.0000	−1.6150	0.0665	0.0000	−2.0133	0.0936	0.0000
x_1	0.0105	0.0009	0.0000	0.0169	0.0015	0.0000	0.0286	0.0020	0.0000
x_2	−0.2613	0.0592	0.0000	−0.5202	0.1052	0.0000	0.0677	0.1242	0.5853
x_3	−0.0024	0.0579	0.9656	0.1526	0.0925	0.0992	−0.0296	0.1325	0.8232
x_4	0.2456	0.0927	0.0081	0.5673	0.1336	0.0000	0.2633	0.1856	0.1559
x_5	−0.4761	0.1711	0.0053	−0.4097	0.2965	0.1671	−0.0200	0.2797	0.9428
x_6	−0.9295	0.0659	0.0000	−0.8913	0.1132	0.0000	−1.2872	0.1602	0.0000
x_7	−0.8464	0.0756	0.0000	−0.8048	0.1291	0.0000	−1.2988	0.2044	0.0000
x_8	−0.9161	0.1456	0.0000	−1.1461	0.2753	0.0000	−1.6853	0.4318	0.0000
		ρ_1			ρ_2			ρ_3	
Intercept	−0.7403	0.0182	0.0000	−1.6842	0.0502	0.0000	−12.2900	0.0989	0.0000
x_1	0.0052	0.0004	0.0000	0.0072	0.0011	0.0000	−0.0026	0.0024	0.2759
x_2	−0.0563	0.0256	0.0277	−0.1685	0.0725	0.0201	−0.1374	0.1396	0.3252
x_3	0.0043	0.0270	0.8727	0.4435	0.0654	0.0000	−0.4118	0.1620	0.0110
x_4	−0.0779	0.0535	0.1449	−1.9145	0.2053	0.0000	9.0674	0.2910	0.0000
x_5	0.0484	0.0746	0.5156	−1.3663	0.2326	0.0000	8.7198	0.6505	0.0000
x_6	0.1280	0.0285	0.0000	−1.0798	0.0740	0.0000	9.5204	0.1249	0.0000
x_7	−0.0699	0.0330	0.0344	−1.4868	0.0980	0.0000	8.2593	0.2544	0.0000
x_8	0.0421	0.0610	0.4892	−1.4348	0.1848	0.0000	8.8984	0.3461	0.0000

The parameters were estimated using the maximum likelihood method by employing an R package called Rsolnp (Ghalanos and Theussl, 2015), a standard tool that can generally computes the solutions to nonlinear optimizations. The parameter estimates for IIM with Pareto II, their standard errors, and p-values were also listed in the following Tables 7 and 8. The tables read that the most covariates were significant at level 0.05. Particularly, larger age tended to increase the number of claims and there was a higher rate of claim occurrence in the general health insurance than in serious illness insurance, while the latter tended to cause larger reporting and S-delays.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 8

ESTIMATED PARAMETERS FOR CLAIM PAYMENTS: IIM.

$\boldsymbol{\gamma}$	Intercept	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
Estimate	8.7763	0.0105	−0.0942	−2.9001	0.1180	0.0214	−0.0378	0.0290	0.1641	0.2173	0.4046	0.8265
Std. error	0.0134	0.0003	0.0190	0.0233	0.0390	0.0552	0.0218	0.0235	0.0452	0.0257	0.0969	0.2525
p-Value	0.0000	0.0000	0.0000	0.0000	0.0025	0.6976	0.0837	0.2174	0.0002	0.0000	0.0000	0.0010
$\hat{\sigma} = 1.6681$												

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 9
ESTIMATED PARAMETERS: IDM.

$\hat{\lambda}$	$\hat{p}_0, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5$	$\hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3$	$\hat{\gamma}^{h'}, \hat{\sigma}$
0.0579	0.2373, 0.4252, 0.1788, 0.0763, 0.0423, 0.0401	0.5889, 0.3471, 0.0480, 0.0160	8.3794, 8.5778, 8.6416, 8.9645,1.6681

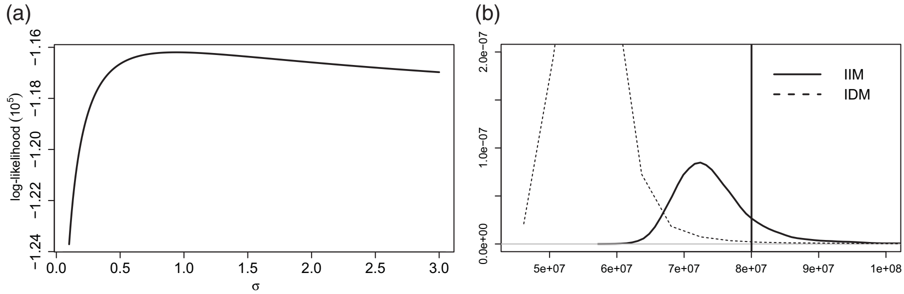

FIGURE 6: The insensitivity of likelihood and the simulated outstanding liabilities. (a) Maximum log-likelihood over σ (b) Estimated densities of $\hat{R}$ in (3.1)

The estimated λ, p, q, γ_v^hs, and σ in IDM are listed in Table 9, computed based on the formulae in (3.15). Note that the direct MLE of σ was originally $\hat{\sigma} = 0.9491$, resulting in an irregular distribution of the payments under which the mean is infinity. However, a careful examination found that the maximum log-likelihood was quite flat when σ took values approximately in [0.9, 2.0]. See the plot to the left of Figure 6 for a visualization of the maximum log-likelihood over σ, which showed that the change in the maximum log-likelihood was small when σ ran up from 0.5 . Therefore, we simply set $\hat{\sigma} = 1.6681$, a value obtained under IIM with Pareto II, which had been shown to fit the data quite well with individual information.

4.4. Loss reserving

Firstly, by substituting their estimates into the computation procedure for the parameters of IIM in Table 1, loss reserving $\hat{R}_m$ was computed in a way displayed in Table 10. $\hat{R}_H$ was also computed in the same manner. The two reservings were listed in Table 11. The table also presented nominal 95% confidence intervals, which were computed with the asymptotic variances (Theorems 3.4 and 3.5) approximated by their empirical versions (substituting the estimates for the parameters).

4.5. Accuracy evaluation

In this real data analysis, the true values of the parameters and thus the the true distribution of claim developments were unknown. Hence, to evaluate the

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 10

COMPUTATION OF LOSS RESERVING $R_m(\hat{\boldsymbol{\theta}}) = R_m^{rbns}(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\gamma}}, \hat{\sigma}) + R_m^{ibnr}(\hat{\boldsymbol{\theta}})$.

RBNS	Settlement delay			Accident	Reporting delay					IBNR
reserve	3	2	1	year	1	2	3	4	5	reserve
590,619	38,906	86,909	464,804	1						
1,141,402	99,090	451,159	591,153	2					1,074,815	1,074,815
2,361,841	418,313	813,786	1,129,742	3				1,173,582	1,239,333	2,412,915
4,786,300	248,457	1,593,137	2,944,706	4			2,459,996	1,289,009	1,364,776	5,113,781
9,565,154		1,250,456	8,314,698	5		6,382,374	2,614,485	1,369,544	1,451,360	11,817,763
8,009,765			8,009,765	6	16,002,858	6,623,571	2,712,566	1,419,821	1,504,758	28,263,574
26,455,081					Total: 75,137,929					48,682,848

TABLE 11

LOSS RESERVINGS, CONFIDENCE INTERVALS (IN MILLIONS, AND TWO-SIDE p-VALUES).

Model	Loss reserving	95% asymptotic confidence interval	Realized	p-Value
IIM	75.1379	(66.6724, 83.6035)	$\geq$ 80.0181	0.2836
IDM	56.6157	(36.4723, 76.7592)	$\geq$ 80.0181	0.0141

prediction capability of IIM, instead of computing the absolute relative error $\frac{|\hat{R}_m - R_m|}{R_m} \times 100$ (as did in the simulation previously), the distribution of outstanding liabilities R at evaluation date June 30 2019 (see (3.1)) was approximated using a Monte Carlo procedure that drew 20000 computerized realizations for every unrealized claims development from the estimated distributions as what follows.

I Monte Carlo under IIM.
Repeat for $J = 1$ to 20,000 to produce Monte Carlo realizations $R_1, R_2, \ldots, R_{20,000}$ as follows:

(i) For each policy (i, k) $(i > n - D^r)$ and every R-delay u satisfying $u > n - i$, first a number N_{iku}^{ibnr} was generated from a Poisson distribution with mean $r_{ik} \exp(\boldsymbol{x}'_{ik}\hat{\boldsymbol{\beta}})\hat{p}_{iku}$, where $\hat{p}_{iku}$ is obtained by plugging $\hat{\boldsymbol{\pi}}$ into p_{iku}, then was a set $(N_{iku0}^{ibnr}, N_{iku1}^{ibnr}, \ldots, N_{ikuD^s}^{ibnr})$ of numbers from a multinomial $M(N_{iku}^{ibnr}; \hat{q}_{ik0}, \hat{q}_{ik1}, \ldots, \hat{q}_{ikD^s})$ if $N_{iku}^{ibnr} > 0$, where $\hat{q}_{ikv} = q_{ikv}(\boldsymbol{x}_{ik}, \hat{\boldsymbol{\rho}})$, and finally was N_{ikuv}^{ibnr} payments from a Pareto II distribution with density $f(y; \hat{\eta}_v, \hat{\sigma})$ with $\log(\eta_v) = \boldsymbol{x}'_{ikv}\hat{\boldsymbol{\gamma}}$ for every $v = 0, 1, \ldots, D^s$.

(ii) For every number N_{iku}^{rbns} (> 0), a set $(N_{iku,n-i-u+1}^{rbns}, \ldots, N_{ikuD^s}^{rbns})$ of numbers was drawn from a multinomial $M(N_{iku}^{rbns}; (\hat{q}_{ik,n-i-u+1}, \ldots, \hat{q}_{ikD^s})/\hat{\hat{Q}}_{ik,n-i-u+1})$, where $\hat{\hat{Q}}_{ik,n-i-u+1}$ was a plug-in estimate as that of $\hat{q}_{ikv}$, and then N_{ikuv}^{rbns} payments from a Pareto II distribution with density $f(y; \hat{\eta}_v, \hat{\sigma})$ with $\log(\eta_v) = \boldsymbol{x}'_{ikv}\hat{\boldsymbol{\gamma}}$ for every $v = n - i - u + 1, \ldots, D^s$.

(iii) Compute R_J^{ibnr}, R_J^{rbns}, and then R_J by means of (3.1)

II Monte Carlo under IDM.
The IBNR and RBNS claims and their developments using the “false” IDM with Pareto II were also generated in the same way described above to produce correspondingly 20,000 Monte Carlo realizations.

The automatically generated densities by an R package from the 20,000 Monte Carlo realizations under IIM and IDM were visualized in part (b) of Figure 6, showing a sharp difference between the distributions of the two predictions. The middle vertical line there represented the real observed outstanding claims (listed also in the fourth column of Table 11). This plot told a much greater accuracy in IIM by including effective individual information than IDM that ignored that useful information. In addition, denoting by

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

TABLE 12

RUN-OFF TRIANGLE FOR CUMULATIVE PAYMENTS AND LOSS RESERVING BY CHAIN-LADDER METHOD.

Accident	Development year						Loss
year	0	1	2	3	4	5	reserving
1	0.00	50,693	1,579,167	3,295,008	4,206,331	5,308,198	–
2	68,834	3,717,531	8,318,154	11,097,503	12,642,863	**15,954,719**	3,311,855
3	1,076,717	8,984,984	15,064,103	19,330,621	**22,630,197**	**28,558,280**	9,227,660
4	2,042,607	13,256,064	23,831,204	**32,196,194**	**37,691,818**	**47,565,360**	23,734,156
5	2,923,321	18,860,600	**35,381,929**	**47,801,339**	**55,960,631**	**70,619,772**	51,759,172
6	2,737,337	**20,097,259**	**37,701,864**	**50,935,594**	**59,629,878**	**75,250,195**	72,512,857
Total: 160.5457×10^6							

R_{rl} realized outstanding liabilities, the two-side "p-values" were computed by $\frac{2}{20{,}000}\min\{\#(R_I < R_{rl}), \#(R_I > R_{rl})\}$. Coincide with the result of Theorem 3.5, the prediction came with a great bias when omitting the individual information which indeed influenced the occurrence of claims and their developments. In this dataset, the reserving $\hat{R}_m$ was larger than $\hat{R}_H$, though the purpose of introducing individual information is to give more accurate rather than higher reservings

4.6. Comparison with chain ladder

To compare with, we also computed a loss reserving by means of the chain-ladder method from the data, as shown in the run-off triangle in Table 12. The estimated developments of cumulative payments by chain-ladder method were displayed in the bottom right part of the table (the boldfaced). The rightmost column listed the loss reservings of accident years and the last row was the total, which gave almost the twice of the the real loss reserve (the fourth column in Table 11).

5. CONCLUSIONS

This research is mainly motivated by the trend of Big Data applications. In terms of a parametric model for loss reserving which takes into account individual information, this paper studied the asymptotic behaviors of loss reserving with and ignoring individual information. Note that this approach clearly makes tremendous sense in the insurance industry because, in most practice, a portfolio usually contains a huge number of policies. In addition to theoretical induction, this study also examined by simulations the error reduction from introducing individual covariates in prediction of the outstanding liabilities.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

The analyzed model, which is parametric in a statistical context, has achieved its purpose to demonstrate the reduction effects of individual information. However, some researchers may be concerned with the limitation that the model may be subjective and thus question its robustness in practical applications. To address this aspect, a recommended next step would be to study this problem under a nonparametric framework.

Another potential limitation of this model is its critical assumption on the independence between individuals. While this would be true in a majority of real insurance operations, it has also long been recognized that, in many circumstances, a degree of dependence exists among individuals. An example of this is in group insurance, in which individuals from a same group may be dependent in facing the same risks. Significant efforts have been made in areas such as pricing, credibility models, among others, and so on to incorporate the effects of dependence among risks. Thus, studying loss reserving under conditions of dependence would also be worth of further research efforts.

Another potential limitation is that in the model we discussed, the same covariates $\boldsymbol{x}$ are shared by the distributions of R-delays, S-delays, and the payments, when, in fact, they may be affected by different sets of covariates. While covariates determinations are basically the business of the mechanism governing the producing of data, model/variable selection procedures may help.

In the insurance industries, loss reserving is generally periodically reviewed, and it would be helpful but challenging to analyze the problems of loss reserving in terms of continuous time stochastic processes in a point of view of mathematics—like the framework proposed by Heras *et al.* (2018)—but leave the distribution of those stochastic processes fully or at least partly unknown.

ACKNOWLEDGEMENT

This work was supported by NSFC (71771089), the Shanghai Philosophy and Social Science Foundation (2015BGL001), and the National Social Science Foundation Key Program of China (17ZDA091).

REFERENCES

ARJAS, E. (1989) The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin: The Journal of the IAA*, **19**(2), 139–152.

AVANZI, B., WONG, B. and YANG, X. (2016) A micro-level claim count model with overdispersion and reporting delays. *Insurance: Mathematics and Economics*, **71**, 1–14.

BADESCU, A.L., LIN, X.S. and TANG, D. (2016) A marked Cox model for the number of IBNR claims: Theory. *Insurance: Mathematics and Economics*, **69**, 29–37.

BOJ, E. and COSTA, T. (2018) Logistic classification for new policyholders taking into account prediction error. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 161–165. Springer.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

CREVECOEUR, J., ANTONIO, K. and VERBELEN, R. (2019) Modeling the number of hidden events subject to observation delay. *European Journal of Operational Research*, **277**(3), 930–944.

DENUIT, M., MARÉCHAL, X., PITREBOIS, S. and WALHIN, J.F. (2007) *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. London: John Wiley & Sons.

ENGLAND, P.D. and VERRALL, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8**(3), 443–518.

ENGLAND, P.D. and VERRALL, R.J. (2006) Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, **1**(2), 221–270.

FUNG, T.C., BADESCU, A. and LIN, X.S. (2020) A new class of severity regression models with an application to IBNR prediction. *North American Actuarial Journal*. doi:10.1080/10920277.2020.1729813.

GABRIELLI, A. and WÜTHRICH, V.M. (2018) An individual claims history simulation machine. *Risks*, **6**(2), 29.

GHALANOS, A. and THEUSSL, S. (2015) Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method. R package version 1.16.

GODECHARLE, E. and ANTONIO, K. (2015) Reserving by conditioning on markers of individual claims: A case study using historical simulation. *North American Actuarial Journal*, **19**(4), 273–288.

GOGOL, D. (1993) Using expected loss ratios in reserving. *Insurance: Mathematics and Economics*, **12**(3), 297–299.

HERAS, A., MORENO, I. and VILAR-ZANÓN, J.L. (2018) An application of two-stage quantile regression to insurance ratemaking. *Scandinavian Actuarial Journal*, **2018**(9), 753–769.

HESSELAGER, O. (1995) Modelling of discretized claim numbers in loss reserving. *ASTIN Bulletin: The Journal of the IAA*, **25**(2), 119–135.

HUANG, J., QIU, C. and WU, X. (2015a) Stochastic loss reserving in discrete time: Individual vs. aggregate data models. *Communications in Statistics-Theory and Methods*, **44**(10), 2180–2206.

HUANG, J., QIU, C., WU, X. and ZHOU, X. (2015b) An individual loss reserving model with independent reporting and settlement. *Insurance: Mathematics and Economics*, **64**, 232–245.

HUANG, J., WU, X. and ZHOU, X. (2016) Asymptotic behaviors of stochastic reserving: Aggregate versus individual models. *European Journal of Operational Research*, **249**(2), 657–666.

JEWELL, W.S. (1989) Predicting IBNYR events and delays: I. continuous time. *ASTIN Bulletin: The Journal of the IAA*, **19**(1), 25–55.

JEWELL, W.S. (1990) Predicting IBNYR events and delays II. Discrete time. *ASTIN Bulletin: The Journal of the IAA*, **20**(1), 93–111.

KUO, K. (2019). Deep Triangle: A deep learning approach to loss reserving. *Risks*, **7**(3), 97.

LARSEN, C.R. (2007). An individual claims reserving model. *ASTIN Bulletin: The Journal of the International Actuarial Association*, **37**(01), 113–132.

MACK, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, **23**(2), 213–225.

NORBERG, R. (1993). Prediction of outstanding liabilities in non-life insurance 1. *ASTIN Bulletin: The Journal of the IAA*, **23**(1), 95–115.

NORBERG, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin: The Journal of the IAA*, **29**(1), 5–25.

PIGEON, M., ANTONIO, K. and DENUIT, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin: The Journal of the IAA*, **43**(3), 399–428.

PIGEON, M., ANTONIO, K. and DENUIT, M. (2014). Individual loss reserving using paid-incurred data. *Insurance: Mathematics and Economics*, **58**, 121–131.

VAN DER VAART, A.W. (2000). *Asymptotic Statistics*. New York: Cambridge University Press.

VERBELEN, R., ANTONIO, K., CLAESKENS, G. and CRÈVECOEUR, J. (2017). Predicting daily IBNR claim counts using a regression approach for the occurrence of claims and their reporting delay. Working paper. Available at https://lirias.kuleuven.be/handle/123456789/580750.

VERRALL, R.J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, **26**(1), 91–99.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

VERRALL, R.J. and WÜTHRICH M.V. (2016). Understanding reporting delay in general insurance. *Risks*, **4**(3), 25.
WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, **50**, 1–25.
WÜTHRICH, M.V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, **2018**(6), 465–480.
WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. West Sussex: John Wiley & Sons.
YU X. and HE, R. (2016). Individual claims reserving models based on marked Cox processes (in Chinese). *Chinese Journal of Applied Probability and Statistics*, **32**(2), 201–219.
ZHAO, X. and ZHOU, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, **46**(2), 290–299.

ZHIGAO WANG
Key Laboratory of Advanced Theory and Application
in Statistics and Data Science-MOE
School of Statistics
East China Normal University
Shanghai, China
E-Mail: wangzhigao2015@163.com

XIANYI WU
Key Laboratory of Advanced Theory
and Application in Statistics and Data Science-MOE
School of Statistics
East China Normal University
Shanghai
China
E-Mail: xywu@stat.ecnu.edu.cn

CHUNJUAN QIU (Corresponding author)
Key Laboratory of Advanced Theory
and Application in Statistics and Data Science-MOE
School of Statistics
East China Normal University
Shanghai, China
E-Mail: cjqiu@stat.ecnu.edu.cn

APPENDIX A. PROOFS OF THEOREMS

A.1. Proof of Lemma 2.1

This proof temporarily uses $\mathbb{E}$ to represent expectation given $(r, \boldsymbol{x})$ so that $I = -\mathbb{E}[\frac{\partial^2 l_{E^o}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}]$. Here presented in only the computation of $\mathbb{E}[\frac{\partial^2 l_{E^o}}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'}]$ and the others can similarly be done using (2.3) and Assumption 1.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

By some algebraic computation, it follows that

$$\frac{\partial^2 l_{\boldsymbol{E}^o}}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'} = -\left[\left(\sum_{v=0}^{D_i^s} N_v^s + \sum_{u=0}^{D_i^r} N_u^{rbns}\right)(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}') + \sum_{u=0}^{D_i^r} N_u^{rbns}\left(\sum_{j=n-i-u}^{D^s}\frac{q_j}{\bar{Q}_{n-i-u}}\boldsymbol{\delta}_j \sum_{j=n-i-u}^{D^s}\frac{q_j}{\bar{Q}_{n-i-u}}\boldsymbol{\delta}_j' - \sum_{j=n-i-u}^{D^s}\frac{q_j}{\bar{Q}_{n-i-u}}\boldsymbol{\delta}_j\boldsymbol{\delta}_j'\right)\right]\otimes\boldsymbol{x}\boldsymbol{x}'.$$

Because $\sum_{v=0}^{D_i^s} N_v^s + \sum_{u=0}^{D_i^r} N_u^{rbns}$ is just the number of those reported claims incurred in accident year i:

$$\mathbb{E}\left(\sum_{v=0}^{D_i^s} N_v^s + \sum_{u=0}^{D_i^r} N_u^{rbns}\right) = r\exp(\boldsymbol{x}'\boldsymbol{\beta})P_{D_i^r}.$$

Observe further that $N_u^{rbns} = 0$ for $i \le n - D^s$ and $0 \le u \le n - i - D^s$. Therefore,

$$\mathbb{E}\left[\frac{\partial^2 l_{\boldsymbol{E}^o}}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'}\right] = -r\exp(\boldsymbol{x}'\boldsymbol{\beta})\left[P_{D_i^r}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}') - \sum_{u=(n-i-D^s+1)_+}^{D_i^r} p_u \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\bar{\boldsymbol{q}}_{n-i-u}) - \dfrac{\bar{\boldsymbol{q}}_{n-i-u}\bar{\boldsymbol{q}}_{n-i-u}'}{\bar{Q}_{n-i-u}} \end{pmatrix}\right]\otimes\boldsymbol{x}\boldsymbol{x}'. \tag{A1}$$

Let $v = n - i - u$ and note that

$$\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}' - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\bar{\boldsymbol{q}}_v) - \frac{\bar{\boldsymbol{q}}_v\bar{\boldsymbol{q}}_v'}{\bar{Q}_v} \end{pmatrix} = \begin{pmatrix} \mathrm{diag}(\boldsymbol{q}_v) - \boldsymbol{q}_v\boldsymbol{q}_v' & -\boldsymbol{q}_v\bar{\boldsymbol{q}}_v' \\ -\bar{\boldsymbol{q}}_v\boldsymbol{q}_v' & \frac{Q_v}{\bar{Q}_v}\bar{\boldsymbol{q}}_v\bar{\boldsymbol{q}}_v' \end{pmatrix}.$$

Then, Equation (A1) gives rise to the desired result.

A.2. Proof of Theorem 2.1

By standard theory of MLE, what is needed is just the computation of the information matrix. Let

$$I_{jm} = \sum_{i=1}^{n}\sum_{k=1}^{m_i} I_{jik}, j = 1, 2, 3. \tag{A2}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

Applying the law of large numbers and exchanging the summation orders:

$$I_1 \triangleq \lim_{m\to\infty} \frac{I_{1m}(\boldsymbol{\beta},\boldsymbol{\pi})}{m} = \mathbb{E}\left[\sum_{u=0}^{D^r}\sum_{i=1}^{n-u} \kappa_i p_u \begin{pmatrix} 1 \\ \boldsymbol{\delta}_u - \boldsymbol{p}\end{pmatrix} (1, \boldsymbol{\delta}_u' - \boldsymbol{p}') \otimes (r\exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{x}\boldsymbol{x}')\right] \text{ a.s.},$$

$$I_2 = \mathbb{E}\Bigg[\Bigg(\sum_{i=1}^{n-D^s}\kappa_i \sum_{u=0}^{n-i-D^s} p_u(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}')$$

$$+\sum_{v=0}^{D^s-1}\sum_{i=(n-v-D^r)\vee 1}^{n-v} \kappa_i p_{n-i-v}\begin{pmatrix} \mathrm{diag}(\boldsymbol{q}_v) - \boldsymbol{q}_v\boldsymbol{q}_v' & -\boldsymbol{q}_v\bar{\boldsymbol{q}}_v' \\ -\bar{\boldsymbol{q}}_v\boldsymbol{q}_v' & \frac{Q_v}{\bar{Q}_v}\bar{\boldsymbol{q}}_v\bar{\boldsymbol{q}}_v'\end{pmatrix}\Bigg)\otimes(r\exp(\boldsymbol{x}'\boldsymbol{\beta})\boldsymbol{x}\boldsymbol{x}')\Bigg] \text{ a.s.},$$

where $I_2 \triangleq \lim_{m\to\infty}\frac{I_{2m}(\boldsymbol{\theta})}{m}$ and

$$I_3 \triangleq \lim_{m\to\infty}\frac{I_{3m}(\boldsymbol{\theta})}{m} = -\mathbb{E}\left[r\exp(\boldsymbol{x}'\boldsymbol{\beta})\sum_{v=0}^{D^s}\sum_{i=1}^{n-v}\kappa_i P_{D^r_{i+v}} q_v \frac{\partial^2 \log f(Y;\eta_v,\sigma)}{\partial(\boldsymbol{\gamma}',\sigma)'\partial(\boldsymbol{\gamma}',\sigma)}\right].$$

A.3. Proof of Theorem 3.1

According to (3.1), the loss reserve can be computed as:

$$\mathbb{E}[R|\mathscr{F}] = \mathbb{E}\left[\sum_{i=1}^{n} R_i^{rbns}|\mathscr{F}\right] + \mathbb{E}\left[\sum_{i=n-D^r+1}^{n} R_i^{ibnr}|\mathscr{F}\right].$$

Firstly, under Assumption 1, the RBNS loss reserve is

$$\mathbb{E}\left[\sum_{i=1}^{n} R_i^{rbns}|\mathscr{F}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{k=1}^{m_i}\sum_{l=1}^{N_{ik}} Y_{ikl} I_{\mathscr{A}_i^{rbns}}(U_{ikl}, V_{ikl})|\mathscr{F}\right]$$

$$=\sum_{i=1}^{n}\sum_{u=0}^{D_i^r}\mathbb{E}\left[\sum_{k=1}^{m_i}\sum_{l=1}^{N_{iku}^{rbns}} Y_{ikl}^*|\mathscr{F}\right]$$

$$=\sum_{v=1}^{D^s}\sum_{i=(n-v-D^r+1)\vee 1}^{n-v+1}\sum_{k=1}^{m_i} N_{ik,n-i-v+1}^{rbns}\tilde{\mu}_{ikv},$$

where, given $\boldsymbol{x}_{ik}$, $Y_{ik1}^*, \ldots, Y_{ikN_{iku}^{rbns}}^*$ are i.i.d variables with the same distribution as that of Y_{ik1} given $V_{ik1} > n-i-u$ and the third equality follows from the definition of $\tilde{\mu}_{ikv}$ (see (3.4)) and the substitution $v = n-i-u$.

Second, recalling that N_{iku} means the number of claims with R-delay u, the IBNR loss reserve can be computed by

$$\mathbb{E}\left[\sum_{i=n-D^r+1}^{n} R_i^{ibnr}|\mathscr{F}\right] = \mathbb{E}\left[\sum_{i=n-D^r+1}^{n}\sum_{k=1}^{m_i}\sum_{l=1}^{N_{ik}} Y_{ikl} I_{\mathscr{A}_i^{ibnr}}(U_{ikl}, V_{ikl})|\mathscr{F}\right]$$
$$= \sum_{i=n-D^r+1}^{n}\sum_{k=1}^{m_i}\sum_{u=n-i+1}^{D^r} E\left[\sum_{l=1}^{N_{iku}} Y_{ikl}|\mathscr{F}\right]$$
$$= \sum_{u=1}^{D^r}\sum_{i=n-u+1}^{n}\sum_{k=1}^{m_i} p_{iku} r_{ik} \exp(\boldsymbol{x}_{ik}'\boldsymbol{\beta})\tilde{\mu}_{ik0}.$$

The proof is then complete.

A.4. Proof of Theorem 3.4

For any $u \in \{0, 1, \ldots, D^r\}$, let $g_u = (g_{u1}, g_{u2}, g_{u3})'$ and $\varphi_u = (\varphi_{u1}, \varphi_{u2}, \varphi_{u3})'$ be such that

$$\begin{cases} g_{u1} = \mathbf{0},\ \mathbf{0} \text{ the zero vector of dimension } (D^r+1)p, & \varphi_{u1} = \mathrm{vec}\,(1, (\boldsymbol{\delta}_u - \boldsymbol{p}))\,\tilde{\mu}_0 \otimes \boldsymbol{x}, \\ g_{u2} = \begin{pmatrix} \mathbf{0} \\ \mathrm{diag}(\overline{\boldsymbol{q}}_{n-i-u})\overline{\boldsymbol{\mu}}_{n-i-u} - \tilde{\mu}_{n-i-u+1}\overline{\boldsymbol{q}}_{n-i-u} \end{pmatrix} \otimes \boldsymbol{x}, & \varphi_{u2} = \begin{pmatrix} \mathbf{0} \\ \mathrm{diag}(\overline{\boldsymbol{q}}_0)\overline{\boldsymbol{\mu}}_0 - \tilde{\mu}_0\overline{\boldsymbol{q}}_0 \end{pmatrix} \otimes \boldsymbol{x}, \\ g_{u3} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_{n-i-u+1} \end{pmatrix} \otimes \boldsymbol{x}, & \varphi_{u3} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_0 \end{pmatrix} \otimes \boldsymbol{x}. \end{cases}$$

First note that

$$\sum_{v=1}^{D^s} \frac{\partial O_v^{rbns}(\boldsymbol{\rho}, \boldsymbol{\gamma}, \sigma)}{\partial \boldsymbol{\theta}} = \sum_{u=0}^{D^r}\sum_{i=(n-u-D^s+1)\vee 1}^{n-u}\sum_{k=1}^{m_i} \frac{N_{iku}^{rbns}}{\sum_{s=n-i-u+1}^{D^s} q_{iks}} g_{iku},$$

where g_{iku} stands for the value of g_u computed on individual (i, k), and

$$\frac{\partial O_u^{ibnr}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=n-u+1}^{n}\sum_{k=1}^{m_i} p_{iku} r_{ik} \exp(\boldsymbol{x}_{ik}'\boldsymbol{\beta})\varphi_{iku},$$

in which φ_{iku} also represents φ_u's value computed on individual (i, k). It follows from the law of large numbers that $\frac{1}{m}\frac{\partial R_m}{\partial \boldsymbol{\theta}} \xrightarrow{P} \tilde{g}_0$, where $\tilde{g}_0 = \mathrm{vec}(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3)$. Therefore, by Theorem 2.1:

$$\frac{1}{\sqrt{m}}(R_m(\hat{\boldsymbol{\theta}}) - R_m(\boldsymbol{\theta})) = \frac{1}{m}\frac{\partial R_m}{\partial \boldsymbol{\theta}}\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p\left(\frac{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|}{\sqrt{m}}\right) \xrightarrow{L} N(0, \sigma_I^2).$$

A.5. Proof of Theorem 3.5

The proof proceeds in two parts with the first for the asymptotic bias (3.23) and the other for the asymptotic normality (3.22). Since r appears as the form of $r\exp(\boldsymbol{x}'\boldsymbol{\beta})$ in the following statements, denote by $\eta(\boldsymbol{\beta}) = r\exp(\boldsymbol{x}'\boldsymbol{\beta})$.

Part I: Asymptotic bias.

Recalling the definition of R_m and $\check{R}_H$ (i.e., (3.6) and (3.20), respectively), it holds true that

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

$$\check{R}_H - R_m = \sum_{v=1}^{D^s} \sum_{u=0}^{D_v^r} \sum_{k=1}^{m_{n-u-v+1}} N_{n-u-v+1,ku}^{rbns} (\check{\tilde{\mu}}_v - \tilde{\mu}_{n-u-v+1,kv})$$
$$+ \sum_{u=1}^{D^r} \sum_{i=n-u+1}^{n} \sum_{k=1}^{m_i} (r_{ik} \check{\lambda}_u \check{\tilde{\mu}}_0 - p_{iku} \eta_{ik}(\boldsymbol{\beta}) \tilde{\mu}_{ik0}).$$

Then, the law of large numbers and some simple algebra operations show that

$$\frac{1}{m}(\check{R}_H - R_m) \overset{a.s.}{\to} \Delta = \sum_{v=1}^{D^s} \sum_{u=0}^{D_v^r} \kappa_{n-v-u+1} \mathbb{E}[p_u \eta(\boldsymbol{\beta}) \bar{Q}_{v-1} (\check{\tilde{\mu}}_v - \tilde{\mu}_v)]$$
$$+ \sum_{u=1}^{D^r} \sum_{i=n-u+1}^{n} \kappa_i \mathbb{E}[p_u \eta(\boldsymbol{\beta}) (\check{\tilde{\mu}}_0 - \tilde{\mu}_0)]. \tag{A3}$$

Part II: Asymptotic Normality.

Simple algebra computation gives

$$\frac{1}{\sqrt{m}}(\hat{R}_H - \check{R}_H) = \sum_{v=1}^{D^s} \frac{G_v}{m} \sqrt{m}(\hat{\tilde{\mu}}_v - \check{\tilde{\mu}}_v) + \sqrt{m}(\hat{\tilde{\mu}}_0 - \check{\tilde{\mu}}_0) \sum_{u=1}^{D^r} \frac{r_{[u]}}{m} \hat{\lambda}_u + \check{\tilde{\mu}}_0 \sum_{u=1}^{D^r} \frac{r_{[u]}}{m} \sqrt{m}(\hat{\lambda}_u - \check{\lambda}_u). \tag{A4}$$

Let $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_{D^r})'$, $\hat{\tilde{\boldsymbol{\mu}}} = (\hat{\tilde{\mu}}_0, \hat{\tilde{\mu}}_1, \ldots, \hat{\tilde{\mu}}_{D^s})'$ with $\hat{\lambda}_u$s and $\hat{\tilde{\mu}}_v$s given by (3.15) and Theorem 3.3, respectively, and $\xi_m = \frac{1}{m}(\check{\tilde{\mu}}_0(r_{[1]}, r_{[2]}, \ldots, r_{[D^r]}), \sum_{u=1}^{D^r} r_{[u]} \hat{\lambda}_u, G_1, G_2, \ldots, G_{D^s})'$. Then, the asymptotic distribution of

$$\frac{1}{\sqrt{m}}(\hat{R}_H - \check{R}_H) = (\xi_m'/m)\sqrt{m}\text{vec}(\hat{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}, \hat{\tilde{\boldsymbol{\mu}}} - \check{\tilde{\boldsymbol{\mu}}}) \tag{A5}$$

can be simply deduced from that of $\hat{\boldsymbol{\lambda}}$ and $\hat{\tilde{\boldsymbol{\mu}}}$, where $\check{\boldsymbol{\lambda}} = (\check{\lambda}_1, \ldots, \check{\lambda}_{D^r})'$ and $\check{\tilde{\boldsymbol{\mu}}} = (\check{\tilde{\mu}}_0, \check{\tilde{\mu}}_1, \ldots, \check{\tilde{\mu}}_{D^s})'$. This long and tedious process is addressed step by step in what follows.

Step 1. The asymptotic normality of statistics $\boldsymbol{r} = (r_{(1)}, r_{(2)}, \ldots, r_{(D^r)})'$, $\boldsymbol{N}^r = (\tilde{N}_1^r, \tilde{N}_2^r \ldots, \tilde{N}_{D^r}^r)'$, $\boldsymbol{S}$ and $s_m(\check{\boldsymbol{\gamma}}^h, \check{\sigma})$, where

$$\boldsymbol{S} = (S_0', S_1', \ldots, S_{D^s}')' \text{ with } S_v = (\tilde{N}_v^s, \tilde{\mathcal{N}}_v^r)'. \tag{A6}$$

Recall that the output of the operator vec($*$) stacks the components of the imput $*$ to form a vector or a matrix in vertical direction depending on what the components are. The law of large numbers yields

$$\begin{cases} r_{(u)}/m \overset{a.s.}{\to} \check{r}_{(u)} := \sum_{i=1}^{n-u} \kappa_i \mathbb{E}[r], u = 1, 2, \ldots, D^r, \\ \tilde{N}_u^r/m \overset{a.s.}{\to} \sum_{i=1}^{n-u} \kappa_i \mathbb{E}[p_u \eta(\boldsymbol{\beta})], u = 1, 2, \ldots, D^r, \\ S_v/m \overset{a.s.}{\to} \check{S}_v := \sum_{i=1}^{n-v} \kappa_i \mathbb{E}\left[P_{D_{i+v}^r} \eta(\boldsymbol{\beta})(q_v, \bar{Q}_{v-1})'\right], v = 0, 1, \ldots, D^s. \end{cases} \tag{A7}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

Denote by $(a_u, |_{u=s_1}^{s_2}) = (a_{s_1}, \ldots, a_{s_2})$ and $K_u = \sum_{i=1}^{n-u} \kappa_i$. By Linderberg–Feller's central limit theorem,

$$\sqrt{m}\left(\frac{1}{m}\text{vec}(\boldsymbol{r}, \boldsymbol{N}^r, \mathbf{S}, s_m) - \text{vec}(\check{\boldsymbol{r}}, \check{\boldsymbol{N}}^r, \check{\mathbf{S}}, \mathbf{0})\right) \xrightarrow{L} N(0, \check{C}), \tag{A8}$$

where $s_m = s_m(\check{\boldsymbol{\gamma}}^h, \check{\sigma})$, $\check{\boldsymbol{r}} = \text{vec}(\check{r}_{(u)}, |_{u=1}^{D^r})$, $\check{\boldsymbol{N}}^r = \text{vec}(K_u \mathbb{E}[p_u \eta(\boldsymbol{\beta})], |_{u=1}^{D^r})$, $\check{\boldsymbol{S}} = \text{vec}(\check{S}_v, |_{v=0}^{D^s})$ and

$$\check{C} = \lim_{m\to\infty} \frac{1}{m}\text{Cov}(\text{vec}(\boldsymbol{r}, \boldsymbol{N}^r, \mathbf{S}, s_m)).$$

We now derive $\check{C}$ under Assumption 1 computed by means of the formula:

$$\text{Cov}(\text{vec}(\boldsymbol{r}, \boldsymbol{N}^r, \mathbf{S}, s_m)) = \mathbb{E}[\text{Cov}(\text{vec}(\boldsymbol{r}, \boldsymbol{N}^r, \boldsymbol{S}, s_m)|\boldsymbol{X})] + \text{Cov}(\mathbb{E}[\text{vec}(\boldsymbol{r}, \boldsymbol{N}^r, \mathbf{S}, s_m)|\boldsymbol{X}]),$$

where the $\boldsymbol{X}$ is composed of the observed covariates and risk exposures. Apparently, $\text{Cov}(\boldsymbol{r}, \text{vec}(\boldsymbol{N}^r, \boldsymbol{S}, s_m)|\boldsymbol{X}) = \mathbf{0}$. Then, combining some algebraic computation, the law of large numbers gives

$$F_v^{ns} \triangleq \lim_{m\to\infty} \frac{\mathbb{E}[\text{Cov}(\boldsymbol{N}^r, S_v|\boldsymbol{X})]}{m} = \text{vec}\left(K_{u+v}\mathbb{E}\left[p_u\eta(\boldsymbol{\beta})q_v\varsigma_v'\right], |_{u=1}^{D^r}\right),$$

$$\overline{F}_v^{ns} \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{N}^r|\boldsymbol{X}], \mathbb{E}[S_v|\boldsymbol{X}])}{m} = \text{vec}\left(\sum_{i=1}^{n-u\vee v} \kappa_i \text{Cov}\left(p_u\eta(\boldsymbol{\beta}), \eta(\boldsymbol{\beta})P_{D^r_{i+v}} q_v\varsigma_v'\right), \Bigg|_{u=1}^{D^r}\right),$$

$$F_v^{ss} \triangleq \lim_{m\to\infty} \frac{\mathbb{E}[\text{Cov}(S_v, S_v|\boldsymbol{X})]}{m} = \sum_{u=0}^{D^r_{v+1}} \mathbb{E}\left[K_{u+v}p_u\eta(\boldsymbol{\beta})A_v\right],$$

$$\overline{F}_0^r \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{r}|\boldsymbol{X}])}{m} = (K_{u\vee s}\text{Var}(r))_{D^r\times D^r}, u, s \in \{1, 2, \ldots, D^r\},$$

$$\overline{F}_1^r \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{r}|\boldsymbol{X}], \mathbb{E}[\boldsymbol{N}^r|\boldsymbol{X}])}{m} = (K_{u\vee s}\text{Cov}(r, \eta(\boldsymbol{\beta})p_s))_{D^r\times D^r}, u, s \in \{1, 2, \ldots, D^r\},$$

$$\overline{F}_2^r \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{r}|\boldsymbol{X}], \mathbb{E}[\boldsymbol{S}|\boldsymbol{X}])}{m} = \text{vec}\left(\left(\sum_{i=1}^{n-u\vee v} \kappa_i \text{Cov}\left(r, \eta(\boldsymbol{\beta})P_{D^r_{i+v}} q_v\varsigma_v'\right), \Bigg|_{v=0}^{D^s}\right), \Bigg|_{u=1}^{D^r}\right),$$

$$\overline{F}^n \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{N}^r|\boldsymbol{X}])}{m} = (\overline{F}_{us}^n)_{D^r\times D^r} \text{ and } \overline{F}^{ss} \triangleq \lim_{m\to\infty} \frac{\text{Cov}(\mathbb{E}[\boldsymbol{S}|\boldsymbol{X}])}{m} = (\overline{F}_{tv}^{ss})_{(D^s+1)\times(D^s+1)}, \tag{A9}$$

where

$$\varsigma_v = (1, \tfrac{\bar{Q}_{v-1}}{q_v})', \quad A_v = q_v\begin{pmatrix} 1 & 1 \\ 1 & \frac{\bar{Q}_{v-1}}{q_v} \end{pmatrix}, \quad \overline{F}_{us}^n = K_{u\vee s}\text{Cov}(p_u\eta(\boldsymbol{\beta}), \eta(\boldsymbol{\beta})p_s), u, s \in \{1, 2, \ldots, D^r\},$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

and $\overline{F}^{ss}_{tv} = \sum_{i=1}^{n-t\vee v} \kappa_i \mathrm{Cov}\left(\eta(\boldsymbol{\beta}) P_{D^r_{i+t}} q_t \varsigma_t, \eta(\boldsymbol{\beta}) P_{D^r_{i+v}} q_v \varsigma_v\right)$, $t, v \in \{0, \ldots, D^s\}$. Also,

$$
\begin{aligned}
F^{nd} &\triangleq \lim_{m\to\infty} \frac{\mathbb{E}[\mathrm{Cov}(\boldsymbol{N}^r, s_m|\boldsymbol{X})]}{m} \\
&= \mathrm{vec}\left(\sum_{v=0}^{D^s} K_{u+v} \mathbb{E}\left[p_u \eta(\boldsymbol{\beta}) q_v \frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)} \right], u = 1, 2, \ldots, D^r \right),
\end{aligned}
$$

$$
F^{sd}_v \triangleq \lim_{m\to\infty} \frac{\mathbb{E}[\mathrm{Cov}(S_v, s_m|\boldsymbol{X})]}{m} = \sum_{t=v}^{D^s} \sum_{u=0}^{D^r} K_{u+t} (I_{\{t=v\}}, 1)' \otimes \mathbb{E}\left[p_u \eta(\boldsymbol{\beta}) q_t \frac{\partial \log f(Y;\gamma^h_t,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)} \right],
$$

and

$$
\begin{aligned}
F^{dd} &\triangleq \lim_{m\to\infty} \frac{\mathbb{E}[\mathrm{Cov}(s_m, s_m|\boldsymbol{X})]}{m} \\
&= \sum_{v=0}^{D^s} \sum_{u=0}^{D^r} K_{u+v} \mathbb{E}\left[p_u \eta(\boldsymbol{\beta}) q_v \frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)'} \frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)} \right].
\end{aligned}
$$

Also denote $\overline{F}^r_3 = \lim_{m\to\infty} \frac{\mathrm{Cov}(\mathbb{E}[\boldsymbol{r}|\boldsymbol{X}],\mathbb{E}[s_m|\boldsymbol{X}])}{m}$ and $\overline{F}^{nd} = \lim_{m\to\infty} \frac{\mathrm{Cov}(\mathbb{E}[\boldsymbol{N}^r|\boldsymbol{X}],\mathbb{E}[s_m|\boldsymbol{X}])}{m}$ that is,

$$
\overline{F}^r_3 = \mathrm{vec}\left(\sum_{v=0}^{D^s} \sum_{i=1}^{n-u\vee v} \kappa_i \mathrm{Cov}\left(r, \eta(\boldsymbol{\beta}) P_{D^r_{i+v}} q_v \mathbb{E}\left[\frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)'} \middle| \boldsymbol{x} \right] \right), \Bigg|_{u=1}^{D^r} \right),
$$

and

$$
\overline{F}^{nd} = \mathrm{vec}\left(\sum_{v=0}^{D^s} \sum_{i=1}^{n-u\vee v} \kappa_i \mathrm{Cov}\left(p_u \eta(\boldsymbol{\beta}), \eta(\boldsymbol{\beta}) P_{D^r_{i+v}} q_v \mathbb{E}\left[\frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)'} \middle| \boldsymbol{x} \right] \right), \Bigg|_{u=1}^{D^r} \right).
$$

Besides,

$$
\begin{aligned}
\overline{F}^{sd}_t &\triangleq \lim_{m\to\infty} \frac{\mathrm{Cov}(\mathbb{E}[S_t|\boldsymbol{X}], \mathbb{E}[s_m|\boldsymbol{X}])}{m} \\
&= \sum_{v=0}^{D^s} \sum_{i=1}^{n-t\vee v} \kappa_i \mathrm{Cov}\left(\eta(\boldsymbol{\beta}) P_{D^r_{i+t}} q_t \varsigma_t, \eta(\boldsymbol{\beta}) P_{D^r_{i+v}} q_v \mathbb{E}\left[\frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)'} \middle| \boldsymbol{x} \right] \right),
\end{aligned}
$$

and

$$
\overline{F}^{dd} \triangleq \lim_{m\to\infty} \frac{\mathrm{Cov}(\mathbb{E}[s_m|\boldsymbol{X}], \mathbb{E}[s_m|\boldsymbol{X}])}{m} = \sum_{i=1}^{n} \kappa_i \mathrm{Var}\left(\eta(\boldsymbol{\beta}) \sum_{v=0}^{D^s_i} P_{D^r_{i+v}} q_v \mathbb{E}\left[\frac{\partial \log f(Y;\gamma^h_v,\sigma)}{\partial(\boldsymbol{\gamma}^{h\prime},\sigma)'} \middle| \boldsymbol{x} \right] \right).
$$

Then, $\overline{F}^{sd} = \mathrm{vec}(\overline{F}^{sd}_0, \overline{F}^{sd}_1, \ldots, \overline{F}^{sd}_{D^s})$.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

The above procedure gives rise to the equality $\check{C} = \check{C}_1 + \check{C}_2$, where

$$\check{C}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\check{N}^r) & F_0^{ns} & F_1^{ns} & \cdots & F_{D^s}^{ns} & F^{nd} \\ \mathbf{0} & F_0^{ns\prime} & F_0^{ss} & \zeta \check{S}_1' & \cdots & \zeta \check{S}_{D^s}' & F_0^{sd} \\ \mathbf{0} & F_1^{ns\prime} & \check{S}_1 \zeta' & F_1^{ss} & \cdots & \zeta \check{S}_{D^s}' & F_1^{sd} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & F_{D^s}^{ns\prime} & \check{S}_{D^s}\zeta' & \check{S}_{D^s}\zeta' & \cdots & F_{D^s}^{ss} & F_{D^s}^{sd} \\ \mathbf{0} & F^{nd\prime} & F_0^{sd\prime} & F_1^{sd\prime} & \cdots & F_{D^s}^{sd\prime} & F^{dd} \end{pmatrix},$$

and

$$\check{C}_2 = \begin{pmatrix} \overline{F}_0^r & \overline{F}_1^r & \overline{F}_2^r & \overline{F}_3^r \\ \overline{F}_1^{r\prime} & \overline{F}^n & \overline{F}^{ns} & \overline{F}^{nd} \\ \overline{F}_2^{r\prime} & \overline{F}^{ns\prime} & \overline{F}^{ss} & \overline{F}^{sd} \\ \overline{F}_3^{r\prime} & \overline{F}^{nd\prime} & \overline{F}^{sd\prime} & \overline{F}^{dd} \end{pmatrix}.$$

Step 2. The asymptotic distribution of $\sqrt{m}\mathrm{vec}(\hat{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}, \hat{\boldsymbol{h}} - \check{\boldsymbol{h}}, \hat{\tilde{\boldsymbol{\mu}}} - \check{\tilde{\boldsymbol{\mu}}})$.

Write $\hat{\tilde{\boldsymbol{\gamma}}} = (\hat{\boldsymbol{\gamma}}^{h\prime}, \hat{\sigma})'$ and correspondingly write $\tilde{\boldsymbol{\gamma}} = (\check{\boldsymbol{\gamma}}^{h\prime}, \check{\sigma})'$, where $(\check{\boldsymbol{\gamma}}^{h\prime}, \check{\sigma})$ are defined in (3.18). Expanding $s_m(\tilde{\boldsymbol{\gamma}})$ at $\hat{\tilde{\boldsymbol{\gamma}}}$ gives rise to

$$\frac{1}{\sqrt{m}} s_m(\tilde{\boldsymbol{\gamma}}) = \left[\int_0^1 \frac{I_{m\gamma}\left(\tilde{\boldsymbol{\gamma}} + t(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}})\right)}{m} \right] dt \cdot \sqrt{m}(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}),$$

where

$$I_{m\gamma}(\boldsymbol{\gamma}^h, \sigma) = -\sum_{v=0}^{D^s} \sum_{l=1}^{\tilde{N}_v^s} \frac{\partial^2 \log f(Y_{vl}; \gamma_v^h, \sigma)}{\partial(\boldsymbol{\gamma}^{h\prime}, \sigma)' \partial(\boldsymbol{\gamma}^{h\prime}, \sigma)}.$$

Under usual regular conditions, the statistic $I_{m\gamma}/m$ almost surely converges to the non-singular matrix:

$$I_\gamma(\boldsymbol{\gamma}^h, \sigma) = -\sum_{v=0}^{D^s} \sum_{i=1}^{n-v} \mathbb{E}\left[\eta(\boldsymbol{\beta}) P_{D_{i+v}^r} q_v \mathbb{E}_{f_x} \left[\frac{\partial^2 \log f(Y; \gamma_v^h, \sigma)}{\partial(\boldsymbol{\gamma}^{h\prime}, \sigma)' \partial(\boldsymbol{\gamma}^{h\prime}, \sigma)} \right] \right]$$

at neighborhood around $\tilde{\boldsymbol{\gamma}}$. Denote by $L_{m0} = \sqrt{m}\mathrm{vec}(\hat{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}, \hat{\boldsymbol{h}} - \check{\boldsymbol{h}}, \hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}})$, where $\hat{\mathbf{h}} = (\hat{h}_0, \hat{h}_1, \ldots, \hat{h}_{D^s})'$ and $\check{\boldsymbol{h}} = (\check{h}_0, \check{h}_1, \ldots, \check{h}_{D^s})'$ with $\hat{h}_v$ given by (3.15) , by which so that $\hat{h}_v = H(S_v/m)$, $v = 0, 1, \ldots, D^s$, where $H(y_1, y_2) = \frac{y_1}{y_2}$. With the partial derivatives of H that is, $\dot{H}_2(y_1, y_2) = (\frac{1}{y_2}, -\frac{y_1}{y_2^2})$, define two block matrices $\Psi = (-\mathrm{diag}(\check{\boldsymbol{\lambda}}), I_{D^r})\mathrm{diag}^{-1}(\check{\boldsymbol{r}})$ and

$$\Phi_0 = \mathrm{diag}\,(\Phi_1, \Phi_{20}) \;\text{ with }\; \Phi_1 = \mathrm{diag}\left(\dot{H}_2(\check{S}_v), \Big|_{v=0}^{D^s} \right) \;\text{ and }\; \Phi_{20} = I_\gamma^{-1}(\boldsymbol{\gamma}^h, \sigma). \qquad \text{(A10)}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

It follows from the delta method that $L_{m0} \xrightarrow{L} N(0, \Sigma_0)$, with $\Sigma_0 = \text{diag}(\Psi, \Phi_0)\check{C}\text{diag}(\Psi', \Phi_0')$. Further, denote by $L_m = \sqrt{m}\text{vec}(\hat{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}, \hat{\boldsymbol{h}} - \check{\boldsymbol{h}}, \hat{\boldsymbol{\mu}} - \check{\boldsymbol{\mu}})$, where $\hat{\boldsymbol{\mu}} = (\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_{D^s})'$, $\check{\boldsymbol{\mu}} = (\check{\mu}_0, \check{\mu}_1, \ldots, \check{\mu}_{D^s})'$ with $\hat{\mu}_v$ given by (3.15). At this point, we need to first derive the asymptotic distribution of L_m, to which the following formula may help

$$\dot{F}_1 = \text{vec}\left(\frac{\partial \mu_v}{\partial(\boldsymbol{\gamma}^{h\prime}, \sigma)}, v = 0, 1, \ldots, D^s\right)_{(\gamma^h, \sigma)' = \tilde{\boldsymbol{\gamma}}},$$

where the partial derivatives:

$$\frac{\partial \mu_v}{\partial \gamma_v^h} = \frac{\partial \mu_v}{\partial \eta_v}\frac{1}{\dot{g}(\eta_v)} \text{ and } \frac{\partial \mu_v}{\partial \gamma_j^h} = 0, j \neq v,$$

with $\mu_v = \mu_v(\gamma_v^h, \sigma)$. It follows again from the delta method that $L_m \xrightarrow{L} N(0, \Sigma)$, with $\Sigma = \text{diag}(\Psi, \Phi)\check{C}\text{diag}(\Psi', \Phi')$, where, recalling Φ_1 in (A10),

$$\Phi = \text{diag}\,(\Phi_1, \Phi_2) \;\text{ with } \Phi_2 = \dot{F}_1 I_\gamma^{-1}(\boldsymbol{\gamma}^h, \sigma). \tag{A11}$$

To compute Σ, divide the matrix $\check{C}$ into a 2×2 block matrix compatible with the block diagonal matrices $\text{diag}(\Psi, \Phi)$ and $\text{diag}(\Psi', \Phi')$ as:

$$\check{C} = \begin{pmatrix} \check{C}_{11} & \check{C}_{12} \\ \check{C}_{12}' & \check{C}_{22} \end{pmatrix}.$$

Hence,

$$\Sigma = \begin{pmatrix} \Psi\check{C}_{11}\Psi' & \Psi\check{C}_{12}\Phi' \\ \Phi\check{C}_{12}'\Psi' & \Phi\check{C}_{22}\Phi' \end{pmatrix}.$$

Set $\boldsymbol{h} = (h_0, h_1, \ldots, h_{D^s})'$, $\boldsymbol{q} = (q_0, q_1, \ldots, q_{D^s})'$, $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_{D^s})'$, and $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_{D^s})'$. Note that $\tilde{\boldsymbol{\mu}}$ is a vector function of $(\boldsymbol{q}, \boldsymbol{\mu})$ and, thus, a function of $(\boldsymbol{h}, \boldsymbol{\mu})$ according to (3.13), so that we denote

$$Q := \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}}|_{\check{\boldsymbol{h}}} = \text{diag}^{-1}(T'\check{\boldsymbol{q}})T'\text{diag}(\check{\boldsymbol{q}}), \tag{A12}$$

where T is a $(D^s + 1) \times (D^s + 1)$ lower triangular matrix with 1 at all nonzero entries, $\check{\boldsymbol{q}} = (\check{q}_0, \check{q}_1, \ldots, \check{q}_{D^s})'$, and

$$W := \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \mathbf{h}}|_{\check{\boldsymbol{\mu}}, \check{\boldsymbol{h}}} = \text{diag}^{-1}(T'\check{\boldsymbol{q}})(T'\text{diag}(\check{\boldsymbol{\mu}}) - \text{diag}(\check{\tilde{\boldsymbol{\mu}}})T')\text{diag}(\check{\boldsymbol{q}}) \begin{pmatrix} \frac{1}{\check{h}_0} & & & \\ -\frac{1}{1-\check{h}_0} & \frac{1}{\check{h}_1} & & \\ \cdots & \cdots & \ddots & \\ -\frac{1}{1-\check{h}_0} & -\frac{1}{1-\check{h}_1} & \cdots & \frac{1}{\check{h}_{D^s}} \end{pmatrix}. \tag{A13}$$

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press

Then it follows again by the delta method that $\sqrt{m}\text{vec}(\hat{\boldsymbol{\lambda}} - \check{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}} - \check{\boldsymbol{\mu}}) \xrightarrow{L} N(0, \check{\Sigma})$ combining the asymptotic results mentioned above, where

$$\check{\Sigma} = \begin{pmatrix} \check{\Sigma}_{11} & \check{\Sigma}_{12} \\ \check{\Sigma}'_{12} & \check{\Sigma}_{22} \end{pmatrix}, \tag{A14}$$

with $\check{\Sigma}_{11} = \Psi \check{C}_{11} \Psi'$, $\check{\Sigma}_{12} = \Psi \check{C}_{12} \Phi'(Q, W)'$ and $\check{\Sigma}_{22} = (Q\Phi_1 + W\Phi_2)\check{C}_{22}(\Phi'_1 Q' + \Phi'_2 W')$.

Step 4. The asymptotic distribution of $\frac{1}{\sqrt{m}}(\hat{R}_H - \check{R}_H)$.

Take a look back to the decomposition (A5). An application of Markov inequality leads to $G_v \xrightarrow{P} \check{G}_v$, $v = 0, 1, 2, \ldots, D^s$. Besides, $r_{[u]}/m$ and $\hat{\lambda}_u$ converge almost surely, of which the limits are given by (3.17). Therefore, $\xi'_m/m \xrightarrow{P} \xi'$ with ξ given by (3.21). Then put the asymptotic distributions in Step 2 and 3 above into the decomposition (A4) gives that

$$\frac{1}{\sqrt{m}}(\hat{R}_H - \check{R}_H) \xrightarrow{L} N(0, \xi' \check{\Sigma} \xi).$$

This completes the proof.

https://doi.org/10.1017/asb.2020.42 Published online by Cambridge University Press