Micro-Level Loss Reserving Models with Applications in Workers Compensation Insurance

Xiaoli Jin
University of Wisconsin-Madison

Abstract

Accurate loss reserves are essential for insurers to maintain adequate capital and to efficiently price their insurance products. Loss reserving for Property & Casualty (P&C) insurance is usually based on macro-level models with aggregate data in a run-off triangle but many limitations exist in these models despite of their simplicity. A run-off triangle is a summary of underlying individual claims data. In recent years, a small set of literature has proposed reserving models that are based on underlying individual claims data, analogous to approaches used in the life insurance industry. They are referred to as “micro-level models”. We demonstrate the application of a micro-level reserving model in a large portfolio of workers compensation insurance provided by a major P&C insurer. A hierarchical model is specified for the individual claim development process, with multiple blocks to model various events (claim occurrence, notification, transaction occurrence and payment amount) during the claim development. The model is estimated with historic data, validated with a hold-out sample, and compared with commonly-used macro-level models. We show that the micro-level model provides a more realistic reserve estimate than that given by the macro-level models, and the estimation error is largely reduced through the use of individual claims data. The micro-level model is more likely to capture the downside potential in reserves and to provide adequate allowance when extreme scenarios occur. We conclude that micro-level models provide valuable alternatives to traditional models for loss reserving.

The data used in this study consist of over five million records and have the nature of “big data”. Our work is similar to data analytics in that we also search for patterns and use these patterns for predictions. It is different from data analytics in the techniques adopted for pattern searching: while big data analytics typically
use data-driven techniques, we employ techniques suggested by our knowledge of
the loss reserving problem built upon a long history of literature.

1 Introduction

In non-life insurance, a loss reserve represents an insurer’s estimate of its outstanding
liabilities for claims that occurred on or before a valuation date. As the largest liability
in insurers’ annual statement, loss reserves have a great impact on insurers’ solvency and
profitability. Hence, accurately estimating the outstanding claims liabilities is extremely
important for insurers. Figure 1 shows the development process of a typical non-life
insurance claim and illustrates why a loss reserve is needed. A claim that occurs at time
$T$ is reported to the insurer at time $W$, then one or several transactions follow to make
payments for the claim until the settlement at time $S$. The gap between occurrence and
reporting, $U$, is referred to as the “reporting delay”, and the gap between reporting and
settlement, $SD$, is referred to as the “settlement delay”. Insurer values the portfolio
periodically. The claim is an incurred-but-not-reported (IBNR) claim at valuation date $\tau_1$; a reported-but-not-settled (RBNS) claim at valuation date $\tau_2$; and a settled claim at
valuation date $\tau_3$. At the first two valuation dates, the claim has a non-zero outstanding
liability that must be covered in the future. So it is necessary for the insurer to estimate
that outstanding liability and set a loss reserve for it.

Figure 1: Development of a typical non-life insurance claim. The claim occurs at time $T$ and
is reported to the insurer at time $W$. Multiple transactions occur at $D_1$, $D_2$ and $D_3$. The claim
is settled at time $S$, and $\tau_1$, $\tau_2$ and $\tau_3$ are three possible valuation dates. Further, $U$ is the
reporting delay and $SD$ is the settlement delay.
Loss reserving for non-life insurance is traditionally based on run-off triangles in which claim counts or amounts are aggregated per accident and development year combination. Table 5 shows the run-off triangle of paid losses for the data used in this paper. An extensive literature has been developed on reserving models with applications to the run-off triangles. They are referred to as the macro-level models in this paper. These models were first developed as deterministic computational algorithms that only provide a point prediction for the outstanding liability. An overview of the deterministic macro-level reserving models is provided by Friedland (2010). Although the deterministic models are still dominant in the current loss reserving practice, they are extended stochastically by researchers usually through distributional assumptions on the aggregate data (cells in the run-off triangle). This extension allows the estimation of reserve uncertainty or even the full predictive distribution of the reserve estimate. Attention has been focused on the estimation of reserve uncertainty, which consists of the parameter (estimation) uncertainty and the process uncertainty. The former is the uncertainty in parameter estimation due to the limited sample size, whereas the latter derives from the intrinsic randomness of the claims development process. England and Verrall (2002) and Wüthrich and Merz (2008) give an overview of the stochastic macro-level models.

The most widely used macro-level model is the chain-ladder method. It was first developed as a deterministic algorithm based on the assumption that claims recorded to date will continue to develop in a similar manner in the future (see Chapter 7 of Friedland (2010) for detailed assumptions and algorithms). A vast literature investigates the statistical basis of the chain-ladder method, with a focus on the distributional assumptions of the aggregate data and the use of generalized linear models (GLMs), see, e.g., over-dispersed Poisson (ODP) model (Renshaw and Verrall (1998)), negative binomial model (Verrall (2000)), Mack’s model (Mack (1993)), and log-normal model (Kremer (1982)). In recent years, the understanding of the chain-ladder technique has been further developed. Kuang et al. (2008, 2011) extends the chain-ladder model with a calendar effect and uses time-series analysis to forecast this effect. Verrall et al. (2010) and Martínez-Miranda et al. (2011, 2012) proposes a double chain-ladder method that simultaneously uses a triangle of paid losses and a triangle of incurred claim counts. Martínez-Miranda et al. (2013) reformulates the triangular data as a histogram and proposes a continuous chain-ladder model through the use of a kernel smoother.

Many paper in the literature discuss issues of the chain-ladder method and other macro-level reserving models. Notably, over-parametrization of the chain-ladder method (Wright (1990)), instable predictions for recent accident years (Bornhuetter and Ferguson
(1972)), problems with the presence of zero or negative cells in run-off triangles (Kunkler (2004)), difficulties in separate assessment of RBNS and IBNR claims (Schnieper (1991); Liu and Verrall (2009)), difficulties in the simultaneous use of incurred and paid claims (Quarg and Mack (2008)), etc. This literature also provides adjustments to address some of the issues, but the adjustments are often suggested in a heuristic fashion and not applied simultaneously.

At the heart of the limitations of macro-level models is the small sample size and the inability to use any information about the individual claims. These issues derive from the inherent nature of the use of aggregate data and thus generally cannot be addressed by any adjustments within the framework of macro-level models. The observed data in a run-off triangle is typically small, leading to a prediction error that could be “disappointingly large” (England and Verrall (2002)). A run-off triangle is essentially a summary of the underlying individual (ideally homogeneous) claims data. If claims are believed to be heterogeneous, then they are often segmented by certain characteristics (usually discrete) and compiled into multiple triangles. In this respect, individual claim level information is used to segment the data before the modeling phase. Nevertheless, under circumstances when the heterogeneity of claims is due to many characteristics (including continuous characteristics), or the characteristics that contribute to the heterogeneity is not clear, or the number of claims in the portfolio is big enough, the segmentation may not be practical and the incorporation of individual claim level information would be desirable in the reserving model. England and Verrall (2002) questioned the continuing use of aggregate data when the underlying extensive micro-level information is available and the computation is feasible. Parodi (2012) points out the misalignment of rate-making and reserving: they both value the same risk but the former is based on individual data whereas the latter is based on aggregate data.

Many limitations of the macro-level models are particularly prominent in long-tail commercial lines. For example, the problem of unstable predictions for recent accident years is more severe for long-tail lines as the observed claims from these years are highly immature. There are higher degree of heterogeneities in commercial line claims than in personal line claims, making the segmentation of data into multiple triangles less practical. For long-tail lines it is particularly important to measure the reserve uncertainty and the downside potential of reserves, so a realistic predictive distribution of the reserve estimate is desirable. The large estimation error from macro-level models may challenge the generation of a realistic predictive distribution.

A small set of academic literature has arisen over the last 20 years that proposes
micro-level stochastic models (also called individual claim level models) for loss reserving. These models use individual claims data as inputs and estimate outstanding liabilities for each individual claim. Norberg (1993, 1999) and Arjas (1989) build a mathematical framework for applying a marked Poisson process in modeling claims development on an individual claim level. Based on this theoretical framework, several articles develop individual claim level loss reserving models and use case studies for illustration, see, e.g., Haastrup and Arias (1997); Larsen (2007); Antonio and Plat (2013), and Antonio et al. (2013). Another stream of literature focuses on predicting the number of IBNR claims with marked Poisson processes. Jewell (1989) presents the theoretical framework. Following this framework, Zhao et al. (2009); Zhao and Zhou (2010) develop models using a semi-parametric specification and use simulated data for illustration.

We are aware of several micro-level studies based on techniques less relevant to the framework of Norberg (1993, 1999). Taylor and McGuire (2004); Taylor et al. (2008) model individual claims with GLMs by incorporating various types of covariates. Mahon (2005) applies a Markov transition matrix to model the ultimate claim sizes on the individual claim level. Rosenlund (2012) proposes a Reserve by Detailed Conditioning method which used individual claims data without any distributional assumptions or likelihood expressions. Taylor and Campbell (2002) tailors a model for workers compensation insurance, which predicts case estimates using GLMs and survival analysis. Parodi (2012) proposes a triangle-free frequency-severity model for individual claims.

Well-specified micro-level models can simultaneously address many limitations of the macro-level models that are mentioned above. Due to the ability to incorporate individual claim level information, micro-level models can efficiently handle heterogeneities in claims data. The large amount of data used in modeling avoids issues of over-parameterization and reduces estimation errors. IBNR and RBNS claims can be easily distinguished by investigating the reporting delay. Paid and incurred data can be used simultaneously in the model (e.g., incorporate case reserves as covariates or through a Bayesian method). With these desirable features, the microlevel model is a valuable alternative to traditional macro-level loss reserving techniques.

Despite the potential benefits of using individual claim level data, the literature on micro-level reserving models is still in its infancy. In sum, we are aware of fewer than 20 research articles on the topic of micro-level reserving, with a focus on mathematical fundamentals. Little attention has been drawn from the insurance practitioners on this topic, as papers that provide detailed and complete implementation of the micro-level models on empirical data are currently lacking in the literature. To our knowledge, Antonio and
Plat (2013) and Antonio et al. (2013) are the only studies that demonstrate such level of detail.

In this study, we demonstrate the application of the micro-level reserving framework proposed by Norberg (1993, 1999) in a large portfolio of workers compensation insurance. A hierarchical model is specified for the individual claim development process, with multiple blocks that model various events (claim occurrence, notification, transaction occurrence and payment amount) during the claim development. The model is estimated with historic data and validated with a hold-out sample. The prediction result is compared with those generated by commonly-used macro-level models. When evaluating the performance of models, we put emphasis on the impact of using micro-level information on the prediction errors and the entire predictive distribution.

Our study adds to the current literature on micro-level models in several ways. First, we apply the micro-level model to workers compensation insurance, which is a commercial line with a very long tail. Investigating micro-level models is particularly valuable for such a line of business, since the limitations of traditional models are more prominent in long-tail commercial lines. Workers compensation is seldom used in studies of micro-level models. While it is used by Taylor and Campbell (2002), their model is particularly tailored for workers compensation and lacks generality. A case study is also lacking in their study. Second, the data used in this study appear to have considerable variations in the claims development patterns over accident years. We extended the model specifications in Antonio and Plat (2013) and show how to model these changing patterns. Since it is common to encounter changing development patterns in reserving practice, this extension largely generalizes the applications of the micro-level model. Third, we decompose the parameter and process uncertainties and investigate the different features of these two uncertainties under macro- and micro-level models. Fourth, we evaluate the impact of micro-level covariate information on the performance of the micro-level model. Lastly, we discuss how one can handle inflation in macro- and micro-level models and compare the micro-level model with an recently proposed extended chain-ladder model that accounts for inflation (Kuang et al. (2008, 2011)).

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 presents the model and constructs the likelihood function. Section 4 shows the estimation results. Section 5 provides the out-of-sample prediction results. Section 6 compares the performance of the proposed micro-level model with some commonly-used macro-level models. Section 7 concludes the paper and points out several directions for future research.
2 Data

We use data from a portfolio of workers compensation insurance provided by a large P&C insurance company in U.S. The data consist of workers compensation claims occurred and reported in 1996-2005. For each claim, a detailed record that tracks the development of the claim up to 12/31/2012, including claim (accident) occurrence date, claim notification date, date and payment amount of each transaction, is provided. Also included in the data are many characteristics about the policy, policy holder, claim, and claimant, e.g., industry of the insured business, demographics of the claimant, and information about the injury.

There are primarily two types of claims in the data, indemnity and medical claims. The former compensates the claimant for the lost wages when off work and the latter reimburses the claimant’s medical expenses. Due to the different natures of the two coverages, indemnity and medical claims appear to have very different development patterns and thus often need to be modeled separately in practice. We focus on medical claims in this study.

Our final sample consists of 1,008,904 medical claims with 5,805,570 transactions. We consider the claim occurrence time, reporting delay, transaction occurrence time and payment amount as the key development information for each claim. Below we present the data by focusing on these different pieces of information.

![Figure 2: Number of claims occurred in each month from January 1996 to December 2005.](image)
**Claim Occurrence Time.** Figure 2 shows number of claims occurred in each month from January 1996 to December 2005. Similar seasonal fluctuations are observed over each year, i.e., highest occurrence in summer and lowest occurrence in winter. This pattern is largely due to the seasonal construction activities. In general, there is a decreasing trend in more recent years. This trend is most likely due to the shrinking business volume of the insurer, but more rigorous analysis cannot be conducted to confirm the source of the trend as the exposure data is not available to us.

**Reporting Delay.** The reporting delay is an important driver of the IBNR reserves. Figure 3 shows the distribution of the reporting delays in months and Table 1 shows the percentage of claims with the reporting delay of 0 day, 1 day,..., 15 days. Half of the claims are reported within a week after the accidents occur, yet, the distribution appears to be highly skewed to the right.

![Figure 3: Distribution of reporting delay. The distribution is censored at 6 months.](image)

**Transaction Occurrence.** We distinguish three types of transactions over the development of a claim. A “Type 1” transaction refers to the settlement of a claim without a payment. A “Type 2” transaction refers to the settlement of a claim with a payment at the same time. A “Type 3” transaction refers to an intermediate payment before the settlement of a claim. The left panel of Figure 4 shows the average number of transactions per claim observed over the development of claims. The occurrence of transactions dramatically increases over the first couple months after notification and then decreases
quickly over time. There are much more Type 3 transactions than the other two types. The right panel of Figure 4 shows the relative frequency of three types of transactions over development months. The proportion of Type 1 transactions is relatively stable over time, the proportion of Type 2 transactions decreases, and the proportion of Type 3 transactions increases over time.

Figure 5 plots the cumulative number of observed transactions per claim over development months. We distinguish claims from several different accident years (AYs) in the figure. On average, each claim has about 4-6 transactions over the observation period. And considerable variation over accident years is observed, i.e., claims from more recent accident years have more transactions\textsuperscript{[1]}.

**Payment Amounts.** Type 2 and 3 transactions come with a positive payment. The distribution of the payment amounts is shown in Figure 6. Payments are heavily right-skewed so the log transformation is performed. A normal fit is provided in the figure. While the log-normal distribution appears to be a reasonable fit to the data, it does not well capture the fat tail. For the application of loss reserving, capturing the tail distribution is particularly important, as the tail corresponds to extremely large payments that may greatly influence the reserve adequacy. Therefore, we also consider (2-parameter) Pareto (also called Lomax) distribution as a candidate for the payment data, as it has a

\textsuperscript{[1]} Accident years 2, 4, 6, 8, 10 are not shown but are consistent with the trend evident in the figure.
Figure 4: Transaction occurrence patterns. Left panel: average number of transactions in each type per claim over development months. Right panel: percentage of transactions in each type over development months.

Figure 5: Cumulative number of transactions per claim over development months by accident year.
heavier tail than the log-normal distribution and it is often used in actuarial applications for data with a very fat tail.

Figure 7 shows the QQ plots for the log-normal fit (left panel) and the Pareto fit (right panel). The figure suggests that the Pareto distribution provides a much better description for the tail whereas the log-normal distribution fits better for small payments. Considering the particular importance of large payments in the loss reserving practice, we select the Pareto distribution over the log-normal distribution[2]

**Initial Case Estimate and Open Claim Indicator.** A case estimate is a claim adjuster’s estimate of the ultimate loss for a reported claim. In the traditional loss reserving practice, the aggregated case estimates are often referred to as the incurred losses. Researchers have been looking for models that can use both paid and incurred losses (see, e.g., Quarg and Mack (2008); Posthuma et al. (2008); Halliwell (2009)). Information on case estimates may add value to micro-level models too. While there are multiple ways to incorporate case estimates in a micro-level model, we use a relatively straightforward specification in this study, i.e., using initial case estimate as a covariate in the model.

[2] Ideally, a distribution with more parameters may be desirable to well describe both the small and large payments. But we put some emphasis on relatively simple and commonly used distributions in this study.
Figure 7: QQ plots (empirical quantiles of the payment data vs. theoretical quantiles of the fitted distribution) for the log-normal fit (left panel) and the Pareto fit (right panel).

of payment amounts. Table 2 shows summary statistics of initial case estimate. The variable has a considerable positive correlation with the payment amounts.

Open claim indicator is a binary variable that takes the value of 1 if the claim is open on the valuation date and takes the value of 0 if the claim is settled on or before the valuation date. Essentially, this indicator gives partial information on the settlement delay and it may depend on a variety of claim- or policy-level characteristics. While directly using the claim- and policy-level characteristics as covariates is more elegant, it also adds considerable complexity to the model (e.g., all the covariates used need to be simulated for the IBNR claims). Hence, we use open claim indicator in this study and consider it as a proxy for the claim- and policy-level information that is not yet used in the model. Table 3 shows the percentage of observations and summary statistics of the payment amount by level of open claim indicator. It appears that the open claims on average have higher and more dispersed payments than the settled claims.

Note that the two covariates are individual claim level characteristics. They can be used in a micro-level loss reserving model but not the traditional macro-level models.
Table 2: Summary statistics of initial case estimate.

<table>
<thead>
<tr>
<th>Mean</th>
<th>St Dev</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
<th>Max</th>
<th>Spearman Corr with Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4,406</td>
<td>14,111</td>
<td>130</td>
<td>400</td>
<td>5,000</td>
<td>10,000</td>
<td>16,704</td>
<td>50,000</td>
<td>793,725</td>
<td>0.1076 (p-value&lt;0.0001)</td>
</tr>
</tbody>
</table>

Table 3: Percentage of data and summary statistics of payment amounts by level of open claim indicator.

<table>
<thead>
<tr>
<th>Open Claim Indicator</th>
<th>% data</th>
<th>Mean</th>
<th>Std</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>84.7%</td>
<td>369</td>
<td>1570</td>
<td>57</td>
<td>124</td>
<td>293</td>
<td>698</td>
<td>1236</td>
<td>4246</td>
</tr>
<tr>
<td>1</td>
<td>15.3%</td>
<td>562</td>
<td>2864</td>
<td>69</td>
<td>154</td>
<td>407</td>
<td>997</td>
<td>1816</td>
<td>6899</td>
</tr>
</tbody>
</table>

3 Model

We present the model in this section. The data consist of over five million records and have the nature of “big data”. Our work is similar to data analytics in that we also search for patterns and use these patterns for predictions. However, it is different from data analytics in the techniques adopted for pattern searching: while data analytics typically use data-driven techniques, we employ techniques suggested by our knowledge of the loss reserving problem built upon a long history of literature.

3.1 Notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Operational definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_i$</td>
<td>Claim occurrence time</td>
<td></td>
</tr>
<tr>
<td>$W_i$</td>
<td>Reporting time</td>
<td></td>
</tr>
<tr>
<td>$S_i$</td>
<td>Settlement time</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>Valuation time (censoring time)</td>
<td></td>
</tr>
<tr>
<td>$D_{ij}$</td>
<td>Time of the $j$th transaction</td>
<td></td>
</tr>
<tr>
<td>$E_{ij}$</td>
<td>Type of the $j$th transaction</td>
<td></td>
</tr>
<tr>
<td>$P_{ij}$</td>
<td>Payment amount in the $j$th transaction</td>
<td></td>
</tr>
<tr>
<td>$U_i$</td>
<td>$W_i - T_i$</td>
<td>Reporting delay</td>
</tr>
<tr>
<td>$SD_i$</td>
<td>$S_i - W_i$</td>
<td>Settlement delay</td>
</tr>
<tr>
<td>$V_{ij}$</td>
<td>$D_{ij} - W_i$</td>
<td>Time of the $j$th transaction since notification</td>
</tr>
<tr>
<td>$J_i$</td>
<td>$\sum_j 1{V_{ij} \leq SD_i}$</td>
<td>Number of transactions</td>
</tr>
<tr>
<td>$J_i^o$</td>
<td>$\sum_j 1{V_{ij} \leq \tau_i}$</td>
<td>Number of observed transactions as of the valuation time</td>
</tr>
</tbody>
</table>

Table 4: Notations for claim $i$.

The occurrence and development of an individual claim is based on a complex process that contains many different events. For an individual claim $i$, we define several notations,
as shown in Table 4.

The claim process is completely described by the claim occurrence time, the reporting delay, the time, type and payment amount of each transaction:

\[ \{T_i, U_i, (V_{ij}, E_{ij}, P_{ij}) : j = 1, 2, ..., J_i \} \]

With respect to a valuation time \( \tau \), claims can be categorized into three groups: incurred-but-not-reported (IBNR) claims, reported-but-not-settled (RBNS) claims, and settled claims. RBNS and settled claims are often referred to as reported claims.

- **IBNR claims**: \( T_i + U_i > \tau \) and \( T_i \leq \tau \). The development process is totally unobserved at time \( \tau \);
- **RBNS claims**: \( T_i + U_i \leq \tau \) and \( T_i + U_i + SD_i > \tau \). But the settlement delay \( SD_i \) is unobserved at \( \tau \), and the claim development process is censored at \( \tau \), i.e., only the partial development \( \{T_i, U_i, (V_{ij}, E_{ij}, P_{ij}) : j = 1, 2, ..., J_i \} \) is observed;
- **Settled claims**: \( T_i + U_i \leq \tau \) and \( T_i + U_i + SD_i \leq \tau \). The full development process is observed.

### 3.2 Marked Poisson Process

Norberg (1993, 1999) formulated a statistical framework for applying the marked Poisson process to non-life insurance loss reserving. According to this framework, we treat the claims process as a marked Poisson process, in which a point is the occurrence time of a claim and the associated mark consists of the reporting delay and the claim development after notification. The development process after notification contains information on each transaction, i.e., the transaction time, transaction type, and payment amount.

With the marked Poisson process, the occurrence of claims follows a non-homogeneous Poisson process with intensity function \( \rho(t) \). Let \( Z \) denote the associated mark, i.e., \( Z_i = (U_i, X_i) \), with \( X_i \) denoting the development process after notification. Then \( Z \) follows a position-dependent distribution \( P_{Z|T} \) and \( P_{Z|T} = P_{U|T} \times P_{X|T,U} \). Claims are random elements in the claim space \( C = [0, \infty) \times Z = [0, \infty) \times [0, \infty) \times X \) with intensity measure:

\[ \rho(dt) \times P_{U|T}(du) \times P_{X|T,U}(dx), \quad (t, u, x) \in C \]

As each disjoint subprocess decomposed from a marked Poisson process is an independent marked Poisson process (see Karr (1991)), we can study the claims reserving
problem in appropriate subclasses by decomposing the claims space $C$. Insurance practitioners often decompose claims into reported claims and IBNR claims. Reported claims belong to the set $C^{rep} = \{(t, u, x) \in C | t + u \leq \tau\}$ while IBNR claims belong to the set $C^{ibnr} = \{(t, u, x) \in C | t \leq \tau, t + u > \tau\}$, each following an independent marked Poisson process. According to Wüthrich and Merz (2008), the process of reported claims has the following intensity measure on $C$:

$$\rho(dt) \times P_{U \mid T}(du) \times P_{X \mid T,U}(dx) \times \mathbb{1}((t, u, x) \in C^{rep}) = \rho(dt) \times P_{U \mid T}(\tau - t) \mathbb{1}(t \in [0, \tau]) \times \frac{P_{U \mid T}(du) \mathbb{1}(u \leq \tau - t)}{P_{U \mid T}(\tau - t)} \times P_{X \mid T,U}(dx),$$

(3)

and the process for IBNR claims has the intensity measure:

$$\rho(dt) \times P_{U \mid T}(du) \times P_{X \mid T,U}(dx) \times \mathbb{1}((t, u, x) \in C^{ibnr}) = \rho(dt) \times (1 - P_{U \mid T}(\tau - t)) \mathbb{1}(t \in [0, \tau]) \times \frac{P_{U \mid T}(du) \mathbb{1}(u > \tau - t)}{1 - P_{U \mid T}(\tau - t)} \times P_{X \mid T,U}(dx).$$

(4)

With respect to the valuation time $\tau$, only the transactions on or before $\tau$ of reported claims can be observed. The observed likelihood of the claims process is thus

$$\prod_{i: t_i + u_i \leq \tau} \rho(t_i)f_{U \mid T}(\tau - t_i) \times \exp \left(- \int_0^\tau \rho(t)F_{U \mid T}(\tau - t)dt\right) \times f_X^{\tau - t_i - u_i}(x_i),$$

(5)

where $f(.)$ denotes a pdf and $F(.)$ denotes a cdf, the superscript in the last term indicates that for a claim occurs at $t_i$ and with reporting delay $u_i$, the development is censored at $\tau - t_i - u_i$ time units after notification.

The development process $X_i$ consists of transaction occurrence time $V_i$, transaction type $E_i$ and payment amount $P_i$. We use a hierarchical specification for the development process and the likelihood is denoted

$$f_{X \mid T,U}(x_i) = f_{V \mid T,U}(v_i) \times f_{E \mid T,U,V}(e_i) \times f_{P \mid T,U,V,E}(p_i).$$

(6)
Combining (5) and (6), the observed likelihood is

$$
\left( \prod_{i : t_i + u_i \leq \tau} \rho(t_i) f_{U|T}(\tau - t_i) \right) \times \exp \left( - \int_0^\tau \rho(t) F_{U|T}(\tau - t) dt \right) \\
\times f_{V|T,U}(v_i) \times f_{E|T,U,V}(e_i) \times f_{P|T,U,V,E}(p_i) \quad (7)
$$

In the original work of Norberg (1993, 1999), the time-varying risk exposure $w(t)$ is taken as the non-homogeneous intensity for the Poisson process. Antonio and Plat (2013) extends the intensity function to allow a time-varying claim occurrence rate $\lambda(t)$, i.e., the intensity function is $w(t)\lambda(t)$, with $w(t)$ known and $\lambda(t)$ a function to be estimated parametrically. We follow the intensity function specification in Antonio and Plat (2013). Because the exposure $w(t)$ for our data is not available, we let $\rho(t) = w(t)\lambda(t)$ and take $\rho(t)$ as the parameter to be estimated. Under the piece-wise constant specification for $\lambda(t)$ proposed in Section 3.3, the lack of information on $w(t)$ has little impact on the reserve calculation for the book of business under consideration. The drawback is that the model can not be used to predict reserves for any new book of business. But when the exposure data is available, which is always the case for insurers, the estimates of $\rho(t)$ can be easily transformed to $\lambda(t)$ and reserves for a new book of business can be calculated.

### 3.3 Distributional Assumptions

The likelihood (7) is built upon five blocks: claim occurrence time, reporting delay, transaction occurrence time, type and payment amount. Below we specify the detailed distributional assumptions for each block. Note that the blocks for transaction occurrence time and type are specified together.

**Reporting Delay.** As a one-time single type event, claim notification can be modeled by distributions from survival analysis. Figure 3 and Table 1 suggest that a large proportion of claims are reported within the first few days after occurrence and the distribution of reporting delay is highly right-skewed. To capture these features, we specify a mixed distribution: a discrete distribution for reporting delays within $n$ days and a Weibull distribution for reporting delays above $n$ days, with $n$ denoting a small integer. The likelihood for the block of reporting delay is given by

$$
\sum_{k=0}^n p_k 1(U = k) + (1 - \sum_{k=0}^n p_k) f_{U|U>n}(u), \quad (8)
$$
where $p_k, k = 0, 1, \ldots, n$ denotes the degenerate probability mass for a reporting delay of $k$ days and $f_U$ denotes the pdf of the Weibull distribution.

**Claim Occurrence Time.** As discussed in Section 3.2, claim occurrence follows a Poisson process with non-homogeneous intensity $\rho(t)$. The occurrence of reported and IBNR claims follow two independent Poisson processes with intensities $\rho(t)F_{U\mid T}(\tau - t)$ and $\rho(t)(1 - F_{U\mid T}(\tau - t))$ respectively. The likelihood for the occurrence of reported claims is

$$L = \left( \prod_{i : t_i + u_i \leq \tau} \rho(t_i)F_U(\tau - t_i) \right) \exp \left( - \int_0^\tau \rho(t)F_U(\tau - t)dt \right). \quad (9)$$

Figure 2 suggests seasonal variations in the claim occurrence. We thus use a piece-wise constant specification on monthly intervals for the non-homogeneous Poisson intensity $\rho(t)$:

$$\rho(t) = \rho_m, \quad m = 1, 2, \ldots, M. \quad (10)$$

Under this piece-wise constant specification, the likelihood (9) turns to

$$L = \prod_{m=1}^M \rho_m^{N_m} \times \prod_{i : t_i + u_i \leq \tau} F_{U\mid T}(\tau - t_i) \times \prod_{m=1}^M \exp \left( -\rho_m \int_{m-1}^m F_{U\mid T}(\tau - t)dt \right). \quad (11)$$

One advantage of the piece-wise constant specification is the resulting relatively simple closed-form expression for the maximum likelihood estimate (MLE) of $\rho$:

$$\rho_m = \frac{N_m}{\int_{m-1}^m F_U(\tau - t)dt}, \quad m = 1, 2, \ldots, M, \quad (12)$$

where $N_m$ is the number of reported claims that occur in month $m$, i.e., $N_m = \sum_{i : t_i + u_i \leq \tau} 1(m-1 \leq t_i < m)$.

**Transaction Occurrence.** Transactions are recurrent events during the claim development and are modeled by the statistical framework of recurrent events (see Cook and Lawless (2007) for details). In particular, we focus on the modeling of the event intensity with hazard rates. As defined earlier, there are three types of transactions under consideration. We thus specify three non-homogeneous hazard rates, one for each type of transactions. They are denoted by $h_1(t)$, $h_2(t)$, and $h_3(t)$ for Type 1, 2, and 3 transactions respectively. Figure 5 indicates the variation in transaction occurrence intensity over accident years. To account for this variation, we incorporate accident year as a covariate in
the hazard model. The hazard rate $h_k(t), k = 1, 2, 3$ is specified by a multiplicative form of a baseline rate $h_k^{(0)}$ and a component that incorporates accident year effects:

$$h_{ik}(t) = h_k^{(0)}(t) \exp(x_i^T \alpha_k), \quad k = 1, 2, 3, \quad (13)$$

where $x_i$ is a vector of indicators for the accident year of claim $i$.

We use a piece-wise constant specification for the baseline hazard rates. With $k = 1, 2, 3$, the baseline rates are:

$$h_k^{(0)}(t) = \begin{cases} 
  h_{k1}^{(0)}, & t \in [0, 1) \\
  \vdots \\
  h_{k6}^{(0)}, & t \in [5, 6) \\
  h_{k7}^{(0)}, & t \in [6, 12) \\
  \vdots \\
  h_{k15}^{(0)}, & t \in [54, 60) \\
  h_{k16}^{(0)}, & t \in [60, 84) \\
  h_{k17}^{(0)}, & t \in [84, \infty) 
\end{cases}$$

Time is in units of months since claim notification. The interval is one month for the first half year, six months for the next 4.5 years, and 24 months thereafter. The intervals are selected according to the observed transaction occurrence patterns from the data.

The likelihood for observed transaction occurrence can be expressed in terms of the hazard rates:

$$L = \prod_{i, t_i + u_i \leq \tau} \prod_{j=1}^{J_i} h_{i1}^{\delta_{ij1}}(v_{ij}) h_{i2}^{\delta_{ij2}}(v_{ij}) h_{i3}^{\delta_{ij3}}(v_{ij}) \times \exp\left(- \int_0^{\tau_i} (h_{i1}(t) + h_{i2}(t) + h_{i3}(t)) dt \right). \quad (14)$$

Here $\delta_{ijk}$ is an indicator that is 1 if the $j$th transaction of claim $i$ is of type $k$, $\tau_i = min(\tau - t_i - u_i, SD_i)$ so $[0, \tau_i]$ gives the observation window for claim $i$. Parameters, including the accident year effects $\alpha_k, k = 1, 2, 3$ and the piece-wise constants in the baseline rates, need to be estimated by maximizing the likelihood.

**Payment Amounts.** Type 2 and 3 transactions come with a positive payment. Figure 6 suggests that the distribution of payment amounts are right-skewed with a fat tail. We use a Pareto distribution to fit the payment data. Several covariate information are taken into account. It is natural to assume payment amounts change over the development of
claims so development year (DY) is incorporated as a covariate. In addition, accident year (AY), open claim indicator, and initial case estimate are also incorporated in the model as they appear to have an impact on the payment amounts suggested by Section 2. The Pareto distribution has two parameters: the scale parameter $\lambda$ and the shape parameter $\theta$. We allow one or two parameters to depend on covariates. Below is the detailed specifications when both parameters are regressed on covariates.

$$P_{ij} \sim \text{Pareto}(\lambda_{ij}, \theta_{ij}), \quad \lambda_{ij} = \exp(z_{ij}^T \beta), \quad \theta_{ij} = \exp(z_{ij}^T \gamma),$$

where $z_{ij}$ is the vector of covariates. When only one parameter is used in regression, the shape parameter $\theta$ is held as constant.

The likelihood for the observed payment amounts is

$$L = \prod_{i:t_i+u_i \leq \tau} \prod_{j=1}^{d_i} f_P(p_{ij}; \lambda_{ij}, \theta_{ij}),$$

with $f_P$ denoting the density function of Pareto distribution.

Pareto distribution is often used to fit insurance data in actuarial practice. But Pareto regression with covariates is not widely used yet. When attention is focused on point predictions (usually the mean), it is not straightforward to interpret the regression coefficients as the regression is on scale and shape parameters, and the moments may not even exist. Nevertheless, these problems are of smaller concern in this study when the predictive distribution is of primary interest.

4 Estimation Results

We use claims from accident years 1996-2004 in this study. Data from the latest accident year (2005) is excluded for future research\footnote{Data from accident year 2005 will serve as another hold-out sample to validate the model in our future research.}. We assume a valuation date of 12/31/2004 and divide the data into a training set and a validation set according to this valuation date. The training set includes claims development up to the valuation date and the validation set includes claims development after the valuation date and up to 12/31/2012. Table 5 shows the training set and validation set (shaded cells) when they are compiled to a run-off triangle. The model proposed in Section 3 is fit to the training set. Estimation results are presented below.
### Table 5: Run-off triangle of incremental paid losses for claims from accident years 1996-2004 over development years 1-9. Losses are shown in thousands of US dollars.

<table>
<thead>
<tr>
<th>Accident Year</th>
<th>Development Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td></td>
<td>78,602</td>
<td>62,118</td>
<td>21,322</td>
<td>9,519</td>
<td>5,881</td>
<td>3,765</td>
<td>2,246</td>
<td>2,490</td>
<td>1,255</td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td>109,833</td>
<td>96,049</td>
<td>26,860</td>
<td>13,085</td>
<td>7,664</td>
<td>4,726</td>
<td>3,325</td>
<td>2,385</td>
<td>1,928</td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td>123,324</td>
<td>110,960</td>
<td>33,076</td>
<td>16,071</td>
<td>8,687</td>
<td>5,850</td>
<td>4,638</td>
<td>3,114</td>
<td>2,713</td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td>135,817</td>
<td>112,294</td>
<td>41,713</td>
<td>19,268</td>
<td>10,867</td>
<td>7,023</td>
<td>5,418</td>
<td>4,416</td>
<td>3,897</td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td>97,211</td>
<td>102,032</td>
<td>31,371</td>
<td>16,117</td>
<td>8,940</td>
<td>6,364</td>
<td>4,851</td>
<td>4,109</td>
<td>4,477</td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td>91,069</td>
<td>88,232</td>
<td>28,147</td>
<td>14,394</td>
<td>9,378</td>
<td>5,998</td>
<td>4,628</td>
<td>4,270</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td>70,048</td>
<td>68,680</td>
<td>20,892</td>
<td>11,781</td>
<td>6,284</td>
<td>4,873</td>
<td>4,825</td>
<td>4,094</td>
<td>3,598</td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td>61,466</td>
<td>52,351</td>
<td>16,140</td>
<td>7,774</td>
<td>5,388</td>
<td>4,564</td>
<td>3,394</td>
<td>2,557</td>
<td>2,327</td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td>50,427</td>
<td>46,166</td>
<td>12,876</td>
<td>8,549</td>
<td>5,867</td>
<td>5,327</td>
<td>4,927</td>
<td>4,740</td>
<td>4,303</td>
</tr>
</tbody>
</table>

**Reporting Delay.** A mixed distribution with 8 degenerate probability masses at 0, 1, . . . , 7 days and a Weibull distribution above 7 days is fit to the data. A comparison of the observed and fitted distributions is shown in Figure 8. The specified mixed distribution appears to provide a reasonable fit to the observed reporting delays.

![Figure 8: Fitted Reporting Delay with a mixture of degenerate and Weibull Distributions. The distributions are censored at the 95th percentile of the observed reporting delay data.](image)

**Claim Occurrence Time.** Using the fitted distribution for reporting delays, we optimize the likelihood (9) over the piece-wise Poisson intensities $\rho_m$, $m = 1, 2, \ldots, 108$. The
point estimates are plotted over accident months in Figure 9. The estimates well capture
the seasonal variations and the recent decreasing trend in the data.

Figure 9: MLEs of the piece-wise intensity for the Poisson process.

Transaction Occurrence. Baseline hazard rates and accident year effects are es-
imated by maximizing the likelihood (14). Then estimated hazard rates for each accident
year are computed. Figure 10 shows the estimates of hazard rates $h_k(t)$, $k = 1, 2, 3$ over
development months since notification. For each type of transactions, estimated hazard
rates for AY 1 and 9 are displayed.

The estimated hazard rates well capture the patterns in the actual transaction occur-
rence data (as shown in Figure 4): a dramatic increase in the first couple months after
notification and then a decrease over development time. There appear to be considerable
variations in the hazard rates for different accident years. Particularly, the estimated
hazard rates for Type 3 transactions are much higher for AY 9 than for AY 1, which is
consistent with the patterns shown in Figure 5.

Payment Amounts. Pareto regression is used to fit the payment amounts. Table 6
shows the comparison of models based on AIC. The table indicates that the model with
both parameters depending on covariates provides a better fit than the model with one
parameter depending on covariates. Worth noting is the impact of the two individual claim
level covariates (initial case estimate and open claim indicator) in the model. Table 6 suggests that the use of these two variables improves the fitting.

Table 7 shows the regression coefficients for the two micro-level covariates. They both have a significant impact on the payment amounts. The results indicate that, on average, the payment amount in each transaction for a claim that is open at the valuation time is 43% higher than that for a claim that is settled as of the valuation time; and when initial case estimate increases by 10%, payment amount increases by 1.33%\footnote{It is not straightforward to interpret the coefficients for regressions on the scale and shape parameters. To be consistent with the traditional way of coefficients interpretations, we still interpret the coefficients in terms of the impact of the covariates on the average payment amount.}.

Table 6: Comparison of the goodness-of-fit for different models.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Regression on Parameters</th>
<th>Micro-Level Covariates</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto</td>
<td>scale</td>
<td>Yes</td>
<td>62,748,808</td>
</tr>
<tr>
<td>Pareto</td>
<td>scale, shape</td>
<td>Yes</td>
<td>62,704,259</td>
</tr>
<tr>
<td>Pareto</td>
<td>scale, shape</td>
<td>No</td>
<td>62,791,910</td>
</tr>
</tbody>
</table>

We also assess the goodness-of-fit by probability integral transform. For an observed payment amount $p_{ij}$, the probability integral transform of the Pareto model is...
Table 7: Pareto regression coefficients for the two micro-level covariates.

<table>
<thead>
<tr>
<th>covariate</th>
<th>scale parameter</th>
<th>shape parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>Open claim indicator</td>
<td>-0.0588</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>log(initial case estimate)</td>
<td>-0.0075</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>-0.4182</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>-0.1457</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

$F(p_{ij}; \lambda_{ij}, \theta_{ij})$, with $F(.)$ denoting the cdf of the Pareto distribution. If the model provides a good fit to the data, then the distribution of the transformed data is uniform on $[0,1]$. Figure 11 compares the distribution of the probability integral transformed payments with the uniform distribution on $[0,1]$. The plot suggests that the Pareto model provides a reasonable fit to the data except that the fitting on the lower percentiles is poorer. A more complicated fat-tail distribution with more parameters might be desirable to improve the fitting in the future.

Figure 11: Probability Integral Transform Plot for the Pareto Model. The x axis is the theoretical percentiles of the uniform distribution on $[0,1]$ and the y axis is the empirical percentiles of the probability integral transformed payments.

## 5 Out-of-Sample Validation

The calibrated micro-level model is used to predict the development of each open claim after the valuation date. Then the predicted outstanding liabilities for open claims are aggregated to obtain a reserve estimate for the book of business. From an actuarial
perspective, this is the loss reserving process. From a predictive modeling perspective, this is the out-of-sample validation for the proposed micro-level model.

5.1 Prediction Routine

Recall that claims are often categorized into IBNR, RBNS and settled claims. For each IBNR claim, we predict the claim occurrence time, reporting delay, and the time, type, and payment amount of each transaction. For each RBNS claim, we predict the time, type and payment amount of each transaction after the valuation time. The prediction routine is described step-by-step as follows.

Step 1. Predict the number of IBNR claims and their occurrence times.

As discussed in Section 3.2, the occurrence of IBNR claims follows a non-homogeneous Poisson process with intensity $\rho(t)(1 - F_{U|T}(\tau - t))$. With a monthly piece-wise constant specification of $\rho(t)$, the number of IBNR claims that occur in month $m$ follows a Poisson distribution:

$$N_{m}^{ibnr} \sim \text{Poisson} \left( \rho_{m} \int_{m-1}^{m} (1 - F_{U|t}(\tau - t)) dt \right).$$

(17)

The parameter of the Poisson distribution can be evaluated by using the estimated distribution of the reporting delay and the MLEs of $\rho_{m}$. Given the simulated number of IBNR claims in a month, the occurrence times of these IBNR claims are simulated from a uniform distribution in that month.

Step 2. Predict the reporting delay for each IBNR claim.

For an IBNR claim that occurs at time $t$, the reporting delay is drawn from the conditional distribution $\Pr(U \leq u|U > \tau - t)$. Using the estimated mixture distribution of the reporting delay, the above conditional distribution is evaluated numerically and the reporting delay for each IBNR claim is simulated by inverting the conditional distribution.

Step 3. Predict the initial case estimate for each IBNR claim.

Ideally, the initial case estimate for each IBNR claim should be simulated from the distribution of initial case estimates. But the distribution of the observed initial case estimates appears to be hard to estimate by a relatively simple distribution. Alternatively, we assume the initial case estimate of each IBNR claim equal to the mean of the observed initial case estimates. This approximation will reduce the process uncertainty in the reserve estimate. But the impact is very small as the number of IBNR claims is small in
the data.

**Step 4. Predict the occurrence times of future transactions.**
Under the survival model for transactions, the distribution of the transaction occurrence time $V_{ij}$ for claim $i$ is given by

$$
Pr(V_{ij} \leq v) = 1 - \exp \left( - \int_0^v (h_{i1}(t) + h_{i2}(t) + h_{i3}(t))dt \right).
$$

(18)

The censoring time, denoted $c_i$, is equal to $\tau - t_i - u_i$ for RBNS claims and 0 for IBNR claims. The next transaction can occur at any time $v > c_i$. To simulate the time of the first transaction after the censoring time we need to invert the conditional probability

$$
Pr(V_{ij} \leq v|V_{ij} > c_i), \quad j = J_i^o + 1.
$$

(19)

Given the time of the prior transaction after the censoring time, $v_{ij-1}$, the time of the next transaction $V_{ij}$ is simulated by inverting the following conditional probability

$$
Pr(V_{ij} \leq v|V_{ij} > v_{ij-1}), \quad j > J_i^o + 1.
$$

(20)

Under the piece-wise constant specification of the baseline hazard rates, the probability inversion can be conducted by either a numerical routine or closed-form expressions.

**Step 5. Predict the types of future transactions.**
Given the simulated time of a transaction, $v_{ij}$, the type of the transaction is simulated from a discrete distribution with probability mass function:

$$
Pr(E_{ij} = k) = h_{ik}(v_{ij})/\sum_{k=1}^3 h_{ik}, \quad k = 1, 2, 3, \quad j > J_i^o.
$$

(21)

**Step 6. Predict the payment amounts of future transactions.**
Given the simulated time and type of the transaction, the payment amount can be simulated. If the transaction is of Type 1, then the payment amount is 0. Otherwise the payment amount is drawn from the estimated Pareto distribution, i.e., $P_{ij} \sim \text{Pareto}(\lambda_{ij}, \theta_{ij})$, $j > J_i^o$, with the estimated distribution parameters.

**Step 7. Determine stop or continue.**
If the simulated transaction is of Type 1 or 2, then the claim is settled and the prediction routine stops. Otherwise another transaction is simulated.

5.2 Prediction Results

With respect to the valuation date 12/31/2004, we use the estimated micro-level model to predict the future development of each open claim and then aggregate the predicted outstanding liabilities to obtain the reserve estimate for the portfolio. We are not only interested in the point prediction but also the entire predictive distribution of the reserve estimate. The prediction routine is repeated 2000 times to generate the predictive distribution. The prediction result is then compared with the actual total outstanding liability given by the validation set.

The left panel of Figure 12 shows the predictive distribution of the reserve estimate from the micro-level model, based on 2000 simulations. The actual total outstanding liability is given by the vertical dashed line. The predictive distribution of the reserve estimate is heavily right-skewed with the actual outstanding liability falling around the 75th percentile of the distribution. We also demonstrate the prediction results for two important quantities: the number of IBNR claims (middle panel) and number of future transactions (right panel). The figure suggests that the predictive distributions given by the micro-level model are generally realistic except that the actual number of IBNR claims is in the left tail of the distribution.

Note that the prediction is up to development year 9 (DY9), the longest development period that we have data for the most recent accident year. Precisely, the reserves we obtain is an estimate of the future liabilities to be paid within DY9, rather than an estimate of the total outstanding liabilities, since claims may continue to develop after DY9. The difference is small though, because very few claims remain open after DY9. In traditional chain-ladder type models, a tail-factor based on judgment or external data is selected in the heuristic fashion to calculate the outstanding liability beyond the latest available DY. With the micro-level model, predictions beyond the latest available DY can be performed in a more rigorous fashion. For the case study under consideration, it is feasible to predict claims development beyond DY9 through some adjustments to the model, e.g., specifying a more parametric form for the time-varying Poisson intensity and hazard rates. Although attention is focused on the prediction within DY9 in this study, an investigation of modeling beyond that period is of interest in the future.
Figure 12: Predictive distribution of the total reserve (left), number of IBNR claims (middle) and number of future transactions (right) generated by the micro-level model (based on 2000 simulations). The vertical dashed line indicates the actual quantity observed in the data. The distribution of the total reserve is censored at its 95th percentile.

5.3 The Role of Micro-Level Covariates

Two individual claim level covariates, initial case estimate and open claim indicator, are used in the micro-level model. Section 4 shows that the use of them improves the goodness-of-fit of the in-sample result. Now we investigate their impact on the prediction results.

Figure 13 compares the predictive distributions of the reserve estimate when modeling with and without initial case estimate and open claim indicator. In the presence of the two micro-level covariates, the distribution shifts to the right and the tail is fatter. Basically, open claim indicator captures the difference in the payment severity between an open and a settled claim, i.e., an open claim has a higher transaction-level payment amount on average, and thus shifts the distribution to the right. The use of initial case estimate allows variation in payment amounts among individual claims, which increases the dispersion of the distribution, especially the heaviness of the tail. The result suggests that the two micro-level covariates play an important role in generating a realistic reserve estimate.
Figure 13: Predictive distributions of the total reserve generated by the micro-level model. The solid line gives the distribution when initial case estimate and open claim indicator are used in the model. The dashed line gives the distribution when initial case estimate and open claim indicator are not used.

6 Comparison with Macro-Level Models

6.1 Comparison with Chain-Ladder Over-Dispersed Poisson Model

To further assess the performance of the micro-level model, we compare the prediction results to those generated by a commonly-used macro-level model: stochastic chain-ladder method with Over-Dispersed Poisson (ODP) assumption. See Renshaw and Verrall (1998) for details of the model. With $C_{ij}, i = 1, 2, \ldots, 9, j = 1, \ldots, 10 - i$, denoting cell $(i, j)$ in the incremental paid loss triangle (the upper triangle in Table 5), the ODP chain-ladder method is specified by

$$C_{ij}/\phi \sim \text{Poisson}(m_{ij}/\phi), \quad \log(m_{ij}) = c + \alpha_i + \beta_j.$$  \hspace{1cm} (22)

To obtain the predictive distribution of the reserve estimate, we adopt the bootstrapping algorithm in England and Verrall (2002). The predictive distribution based on 2000 simulations is shown in the left panel of Figure 14. The predictive distribution generated by the micro-level model is re-plotted in the right panel for comparison. With the chain-ladder ODP model, the actual outstanding liability is to the right of the entire predictive distribution, suggesting that the model underestimates the outstanding liability by a considerable amount. We conclude that the micro-level model gives more realistic prediction
Figure 14: Predictive distribution of the total reserve generated by the chain-ladder over-dispersed Poisson model (left) and the micro-level model (right). Distributions are based on 2000 simulations. The dash line indicates the actual outstanding liability observed in the data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Std Dev (9,833)</th>
<th>median</th>
<th>Var(75)</th>
<th>Var(90)</th>
<th>Var(95)</th>
<th>Var(99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-Level</td>
<td>235,493</td>
<td>46,367</td>
<td>228,970</td>
<td>235,843</td>
<td>246,821</td>
<td>259,842</td>
<td>338,456</td>
</tr>
<tr>
<td>CL ODP</td>
<td>190,070</td>
<td>8,732</td>
<td>190,225</td>
<td>195,879</td>
<td>200,986</td>
<td>204,451</td>
<td>209,381</td>
</tr>
</tbody>
</table>

Table 8: Comparison of the predictive distributions from the chain-ladder ODP and the micro-level model. The number in parentheses gives the standard deviation for distribution of the micro-level model when the top 5th percent of the simulated data are excluded.

than the chain-ladder ODP model.

Table 8 compares the mean, standard deviation, and risk measures of the predictive distributions from the micro-level model and the chain-ladder ODP method. The standard deviation of the predictive distribution estimates the reserve uncertainty. It appears that the micro-level reserve estimate has a much larger uncertainty than the chain-ladder ODP estimate. However, as the distribution from the micro-level model is highly skewed, standard deviation does not provide much information on its dispersion. When excluding the top 5% extreme values, the standard deviation decreases to 8,833, which is comparable to the standard deviation of the chain-ladder ODP model. The comparison of the risk measures suggests that the micro-level model is more likely to capture the downside potential of reserves and to provide adequate allowance when the extreme adverse scenarios occur.
6.2 Parameter Uncertainty and Process Uncertainty

The uncertainty of predictions can be decomposed into two components: parameter uncertainty and process uncertainty (see, e.g., England and Verrall (2002); Taylor (2014)). The former comes from the uncertainty in the estimation of parameters of the reserving model due to the limited sample size, whereas the latter comes from the intrinsic randomness of the claims development in the future. The predictive distributions in Figure 14 includes both parameter and process uncertainties. For the chain-ladder ODP model, parameter uncertainty is introduced by re-sampling the observed claims development through the bootstrapping method and the process uncertainty is introduced by simulating the future development from the estimated ODP distribution. While this bootstrapping algorithm provides a standard way to include both parameter and process uncertainties, it is very computationally intensive for the micro-level model. To reduce the computational load, we include parameter uncertainty for the micro-level model through the asymptotic normal distribution of the MLEs of the parameters. That is, in each repetition of the prediction routine, we draw a vector of parameters from the asymptotic multivariate normal distribution of the MLEs for each block of the micro-level model.

Table 8 suggests the prediction uncertainties of the chain-ladder ODP and micro-level models are comparable. Comparing each component of the prediction uncertainty is of great interest too. Below we eliminate the process uncertainty and then compare the parameter uncertainty of the two models. For the chain-ladder ODP model, we eliminate the process uncertainty by using the expected values of the future cells in each bootstrapping step. For the micro-level model, however, it is not straightforward to get the expected values in the reserve estimate with the presence of multiple blocks, especially the survival model blocks. For simplicity, we take expected values only for the block of payment amounts, which eliminates most of the process uncertainty since the fat-tailed Pareto model is the largest source of process uncertainty.

Figure 15 compares the predictive distributions with and without process uncertainties for the chain-ladder ODP model (left panel) and the micro-level model (right panel). The figure suggests that while the parameter uncertainty is the primary contributor to the prediction error of the chain-ladder ODP model, it is only a very small proportion of the prediction error given by the micro-level model. The parameter uncertainty of the micro-level model is much smaller than that of the chain-ladder ODP model. This is primarily due to the amount of data used in the estimation, i.e., the chain-ladder ODP model uses 45 cells in the run-off triangle to estimate parameters while the micro-level model uses millions of data records. The process uncertainty, especially that from the distribution
of the payment amounts, has a great impact on the shape of the predictive distribution from the micro-level model. This implies that great caution is in need when selecting a model for the payment amounts under the hierarchical specification of the current model.

Figure 15: Predictive distributions of the reserve estimate with parameter and process uncertainties (solid line) and with only parameter uncertainty (dashed line). Distributions are based on 2000 simulations. The left panel shows results from the chain-ladder ODP model and right panel shows results from the micro-level model. Without process uncertainties, the standard deviation is 6,923 for the chain-ladder ODP model and 1,704 for the micro-level model.

6.3 Comparison with Chain-Ladder Pareto Model

Since the micro-level model uses Pareto distribution for the payment amounts, we also apply a chain-ladder type Pareto model to the run-off triangle for comparison. The model specification is

\[ C_{ij} \sim \text{Pareto}(\lambda_{ij}, \theta_{ij}), \quad E(\lambda_{ij}) = c_1 + \alpha_{1i} + \beta_{1j}, \quad E(\theta_{ij}) = c_2 + \alpha_{2i} + \beta_{2j}. \]  

The specification is similar to that used for the micro-level data, but the two individual claim level covariates, initial case estimate and open claim indicator, cannot be incorporated here for the macro-level model.

The point prediction of the outstanding liability given by the chain-ladder Pareto model is 197,818. It is close to the point prediction given by the micro-level model when
the two micro-level covariates are not used. This result further confirms the importance of the micro-level covariates in generating sufficient reserves.

Note that chain-ladder models with Pareto distributions are seldom seen in the literature. We apply it to the data mainly for the purpose of comparison. Yet, several papers in the literature use Pareto models or consider Pareto models in future research, see, e.g., Venter (2007); Schiegl (2002); Meyers (2009).

6.4 A Note on Inflation

Workers compensation medical claims are often exposed to high medical inflation applicable to the calendar year dimension. In practice, actuaries often discount the claims before modeling and then inflate the output of the model to get a reserve estimate under inflation. For the medical claims under consideration, we experimented by discounting them to 1/1/1996 with the US medical consumer price inflation (CPI) before both the chain-ladder ODP and the micro-level modeling. There appears to be an improvement in the prediction results from both models when compared with the actual outstanding liability (results not reported here), but the the micro-level model still provides a more realistic predictive distribution than the chain-ladder method.

Nevertheless, the selection of inflation rates to discount claims highly depends on judgment or external information and often not realistic enough to reflect the complexity in real-world inflation structure. While medical CPI is a sensible selection for discounting medical claims, it is usually not adequate to capture the claim escalation observed in workers compensation data. And an even bigger problem is that the future medical CPI is actually unknown at the valuation date.

Using the observed data to estimate and predict inflation rates may better capture the inflation structure for the specific portfolio under consideration. Unfortunately, estimating calendar year inflation in traditional models suffers a well-known identification problem (see Kuang et al. (2008, 2011)). Kuang et al. (2008, 2011) also proposes an extended chain-ladder model that provides a solution to the identification problem. Their model extends the chain-ladder log-normal model

\[
\log(C_{ij}) = c + \alpha_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)
\] (24)

to account for calendar year effects (inflation):

\[
\log(C_{ij}) = c + \alpha_i + \beta_j + \gamma_{i+j-1} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).
\] (25)
The model is re-parameterized to achieve unique identification. They continue to propose three time series forecast models (linear trend, level shift, and slope shift of the linear trend) to project the inflation rates after the valuation date.

We apply the extended chain-ladder method to the run-off triangle of the current data. Table 9 shows the point predictions of the outstanding liability generated by the basic and extended log-normal chain-ladder models. Three time series forecast models, as proposed by Kuang et al. (2008, 2011), are used to predict the future inflation rates. It appears that the point prediction from the basic log-normal model is close to that given by the ODP model and also underestimates the outstanding liability. The incorporation of inflation effects does not improve the chain-ladder model’s performance. The lack of improvement may derive from the limited information on the inflation time series that one can extract from the run-off triangle and the relative simplicity of the time series forecast models used here.

The identification problem persists in the micro-level modeling. The model proposed by Kuang et al. (2008, 2011) provides great insights in modeling inflation rates that we can possibly carry on to the micro-level models. The individual claim level development data may provide much finer information on the time series of inflation so a closer modeling and more realistic forecast models for inflation may be obtained. However, the application of the spirit of Kuang et al. (2008, 2011) to micro-level data is far from straightforward and further research is in need.

### 7 Conclusions

We provide a detailed demonstration of the construction and validation of a micro-level reserving model on a large portfolio of workers compensation claims. The model tracks various events during the life-time development of each individual claim. In particular,
we use a hierarchical structure with multiple blocks to model claim occurrence, notification, transaction occurrence and payment amounts. The model follows the mathematical framework proposed by Norberg (1993, 1999) and extends the specifications used by Antonio and Plat (2013) to account for the considerable variations in claim development patterns over accident years.

The model is calibrated with the “past” claims development data up to the valuation date and validated with the hold-out “future” development data. The prediction results are also compared with results from several commonly-used macro-level models. The micro-level model provides a more realistic predictive distribution of the reserve estimate when compared with the macro-level models: the predictive distribution generated by the micro-level model captures the actual outstanding liability at its 75th percentile, whereas the entire distribution of the chain-ladder reserve estimate falls to the left of the actual value. Two micro-level covariates are used and the impact of them on the prediction is evaluated. The use of these covariates helps to handle the heterogeneity in claims and considerably improves the performance of the micro-level model.

We decompose the parameter and process uncertainties of the reserve estimates. The micro-level model generates a much smaller parameter uncertainty than the chain-ladder method, consistent with the different amount of information used in each model. The magnitude of the process uncertainty and the shape of the predictive distribution highly depend on the distributional assumption for the payment amounts under our current micro-level model specification. With the Pareto assumption in this study, the predictive distribution from the micro-level model is highly right-skewed and the process uncertainty is larger than that given by the chain-ladder method. The micro-level model is more likely to capture the downside potential of reserves and to provide adequate allowance when the extreme adverse scenarios occur. This also implies that great caution is in need when selecting assumptions for the micro-level model, especially when the risk measures and the downside potential of reserves are of particular interest.

When claims are discounted to adjust for inflation, the performance of both the chain-ladder and the micro-level models are improved but the latter still outperforms the former. We also apply a more scientific method (extended chain-ladder method proposed by Kuang et al. (2008, 2011)) to account for inflation for the macro-level model but do not observe an improvement in the prediction results.

We show how the micro-level models address many limitations of the traditional models. Predictions for IBNR and RBNS claims are naturally separated. Information on paid and incurred claims are used simultaneously. The use of micro-level development data and
covariate information resolves problems of over-parametrization, reduces the estimation errors and provides more realistic predictions. It is demonstrated that micro-level models provide valuable alternatives to traditional aggregate models, especially when a long-tail commercial line is under consideration.

We plan to apply several extensions to this case study in the near future. First, more rigorous statistical techniques are in need to quantify the performance of different models. Instead of comparing the predictive distribution with one realization of the outstanding liability, techniques such as bootstrapping can be used to obtain a more reliable comparison. Second, more information on the individual claim level can be used as covariates in the micro-level model to better handle the heterogeneity in the data. The currently used covariate, open claim indicator, serves as a proxy for the claim level characteristics that are not yet used in the model. As it considerably improves the performance of the model, it is desirable to test the impact of other claim-level characteristics in the next move. Third, we want to compare the micro-level model with other macro-level models, especially those developed in recent literature. Fourth, Kuang et al. (2008, 2011) sheds light on the way to handle inflation in a micro-level model. Some combination of their spirit and the micro-level data may improve the loss reserving practice under inflation. Lastly, dependence modeling of medical and indemnity claims is of interest, as the occurrence and payments for the two types of claims are likely associated in workers compensation insurance (see Taylor and Campbell (2002)). Associations among claims of multiple coverages may exist in many lines of business, so dependence models may bring great insight in the loss reserving practice.

References


37


