Measuring Operational Excellence in Reserving

Johannes Schneider* and Vincenzo Salipante, Zurich Insurances, Switzerland

Abstract

Motivation: For all insurance companies success depends highly on operational skills. (Large) deviations from estimates of incurred to actual incurred, ie. errors in reserving, increase the risk of bankruptcy in case of unfortunate events as well as to unnecessary tied up resources. Quantifying reserving skills in an easy to understand and accurate manner is a necessity for continuous improvement and operational excellence. To this date, there is no commonly known metric for operational performance in reserving.

Method: We define a new metric called *Reserving Excellence* which touches upon several important aspects of reserving. Most of them can be derived from the time series of the incurred of (closed) claims. It is data-driven in the sense that no parameters must be set by experts, they are derived from the data.

Results: The metric gives a new perspective on operational excellence. The model assumptions could be verified (at least on a qualitative basis) with actual data.

Conclusions: Our key performance indicator for operational excellence of reserving allows to track the impact of initiatives targeting the reserving process as well as to spot problems in the reserving process by considering a single metric and drill-down to individual claims. Still, the metric is a first step and it could be refined further to cover more aspects relevant to reserving.

1 Introduction

Reserving activities are of key interest for insurance companies. Claims reserves should be as accurate as possible, since they are an essential part of overall financial reserves of an insurance corporation. Under-reserving of reported claims might lead to bankruptcy in extreme cases. Over-reserving prevents monetary assets from being invested (more) effectively. Additionally, frequent adjustments of claims should be avoided to maximize operational efficiency. Furthermore, big (and frequent) reserving adjustments should be minimized to reduce uncertainty in the amount of reserves and, thus, also profit of the insurance company. Large reserve strengthening can impact the stock price considerably and it can reduce trust among customers and shareholders.
Therefore, good reserving skills are very important for an insurer. Since the reserving process is far from being trivial, there are several metrics of interest with (complex) interactions. However, management is generally keener on having a single, accurate (aggregate) and reliable metric to save time in analysis and to spot problems at a glance. The overall metric should also allow to give a more-fined grained view of the situation, i.e. it should support drill-down capabilities to exactly determine the root of a problem. This motivates our work. Aside from motivating and defining of our metric, we also provide several cases showing how the metric can be interpreted and how an analysis using drill-down can be done.

In this article, we focus mainly on financial information for analysis, ie. for each claim we consider its incurred from the opening until the closing of the claim. The incurred at a certain time is the sum of the current reserve and total paid at that time. For a perfectly reserved claim, after setting the initial reserves, for every payment the reserves should be reduced by the amount of the payment. Other than that no adjustments should be made to the reserves and the incurred should remain constant (Figure 2). In practice, the reserving process is not accurate resulting in over-reserved and under-reserved claims (Figure 1).

The Reserving Excellence measures among other factors the gap between the incurred over time of (over- and under-reserved) claims and a perfectly reserved claim (Figure 3). It addresses the following question: "How good is the reserving process for reported claims given their reserving history?" Our focus is clearly on assessing operational capabilities. Note, it could be that operational skills are poor, but overall reserves are set correctly, since the errors in reserves of individual claims cancel. This phenomena is rather unlikely for a large number of claims. Aside from the fact that there could be a bias towards over- or under-reserving of single claims, the variance of aggregate claims increases with the number of claims leading to a larger expected reserving error.[1] Therefore, the overall error depends on how well individual claims are reserved, ie. Reserving

*Most work conducted while non-affiliated.

[1] For illustration, assume error in reserves are Gaussian distributed with zero mean. Then the aggregated expected error is still zero, however, the variance of the aggregate is the sum of individual variances of all claims.

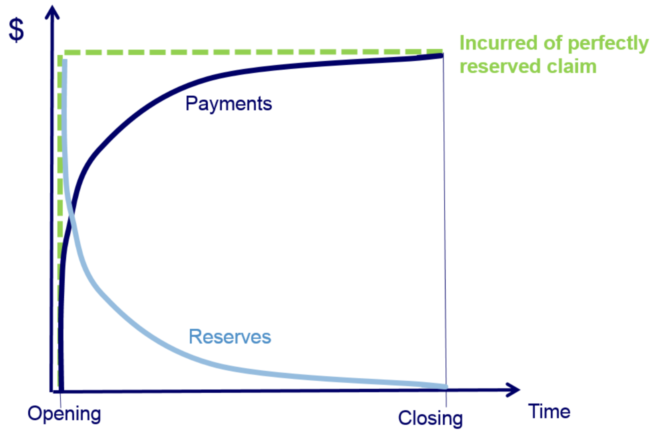

Figure 1: Assuming that the incurred equals payments plus reserves, the incurred of a perfectly reserved claim remains constant.

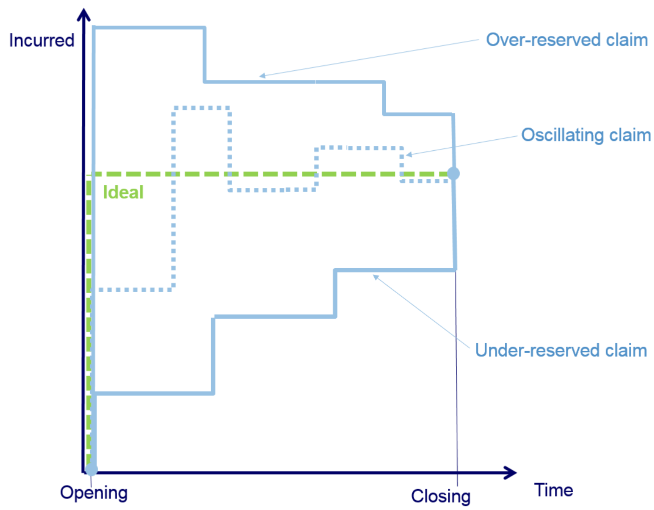

Figure 2: Incurred of over- and under-reserved claims as well as of an ideal claim with adjustments.

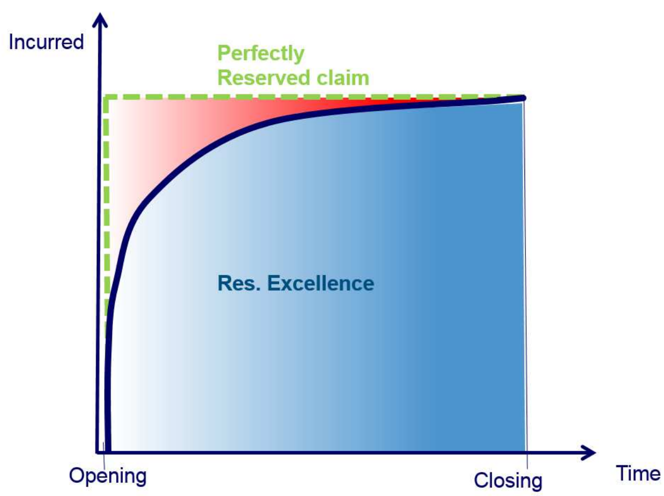

Figure 3: The Reserving Excellence is a value between 0 and 1. For a single claim it depends among other factors on the blue area under the incurred curve. The gap to the optimal claim is shown by the red area. The gradient fill from left to right indicates that initial errors in reserves have less impact on the Reserving Excellence than errors near the closure of the claim.

Excellence. A small error implies more reliable financial forecasting and, therefore, less risk of bankruptcy for an insurer. To what extend the magnitude of errors of individual claims contribute to the measure of Reserving Excellence is a matter of several factors depending on the company and policy (as discussed later). One can argue in favor of absolute or relative errors, ie. a large error for a simple claim (with little overall financial impact) can be considered worse than a small error (with large financial impact) of a large loss or the other way around.

To compute our metric, we proceed in three steps:

1. Determine *reserving difficulty*: We give a rating on how difficult the claim is to reserve. The rating is derived from data. For example, the data might support a statement that (in expectation) a large loss is more difficult to handle than an attritional loss.

2. Judge *reserving performance* for several aspects: We evaluate the reserving history of a (set of closed) claim(s) incorporating company specific goals using different aspects. For example, the severity of adjustments are considered as well as the number of adjustments.

3. Compute *Reserving Excellence* by combining reserving difficulty and the rating of reserving performance. For instance, a claim that is never adjusted (large reserving performance) has larger Reserving Excellence than a claim which reserves are constantly adjusted (low reserving performance). However, for a claim that is difficult to reserve, one might be more tolerant with respect to the number of adjustments: One might say that a complex loss is still handled better than a typical attritional claim, if the complex loss is adjusted twice or thrice and the attritional claim only once.

We discuss all steps top down in later chapters, ie. from a high level management point of view down to precise mathematical models that are required to perform the computations.

1.1 Contributions

1. *Metric(s) for Operational Excellence in Reserving*: We provide the first steps for a single metric for operational excellence. On the way of deriving our metric we discuss many factors that contribute towards the understanding of key performance indicators for claims reserving. These factors by themselves could serve as metrics. We also show factors that are not included in our work and discuss alternatives depending on the strategy of a company.

2. *Data driven approach*: We derive all essential parameters used in our models from collected data. This data driven approach is in contrast to letting parameters be set by experts. Though we do not exclude this possibility, we believe that deriving as many insights as possible from the data has a couple of advantages. First of all, the evaluation of the metric is transparent and can be understood by everyone with very basic and standard mathematical knowledge, whereas parameters set by experts appear as black boxes that must be trusted. This makes not only error detection (in the parameters or model) hard but also limits the analytical drill-down possibilities, since the knowledge needed for explanation is solely "within the expert". Note, that experts are still needed for the derivation of the model. But the model should be more stable and long term than data-dependent parameters. It should also be based on clearly stated assumptions. Second, for parameters based on data, all parameters are automatically adjusted with new data, whereas otherwise experts must adjust them manually (over time). This saves on costs. Third, the parameter adjustments can be more subtle, if they are based on data, since humans will most likely not notice "small" changes, eg. a change of a few percent is likely to be unnoticed though this change can have a major business impact. The gain in accuracy using our data-driven approach makes changes detectable more easily. But, more importantly, changes are discovered earlier, i.e. when major trends start to emerge.

3. *Inclusions of strategic intentions in KPIs:* We also give examples of metrics that reflect forward looking, strategic intentions of a company or operational practices and that are not related to past data. For instance, some companies might prefer frequent adjustments to have accurate reserve estimates whereas others might try to minimize work of claims controllers and explicitly tolerate inaccuracies in estimates. Another example is that a goal could be to minimize the number of adjustments or to prefer conservative reserving, ie. to rate over-reserved claims better than under-reserved claims. In this work, we touch upon a few examples for company specific adjustments.

4. *Drill down capability:* We provide a single metric for operational excellence that is composed of several weighted sub-metrics. These metrics in turn are derived based on the contributions of single claims which are

treated differently based on several characteristics. Our approach is in contrast to traditional metrics that deal primarily with aggregates to compare historical or planned financial (rather than operational) performance with actual figures. Though there might be a correlation between financial performance and operational excellence, there are differences. It is not necessarily the case that good financial performance implies that the reserving process works also well. For instance, assume that the errors in reserving of the individual claims cancel out, ie. deviations in incurred due to over-reserving equal the deviation due to under-reserving. Then, overall judging only from a financial point of view the reserving process seems excellent, ie. perfectly accurate, but from an operational point it is not, since potentially many claim have erroneous reserves. Our key metric uses the financial history of each claim to estimate the deviation from an optimal reserving process which sets the reserves correctly at all times. Such an approach enables drill down, eg. to extract the best or worst reserved claims. Furthermore, our metrics treat different aspects relevant to operational performance separately, which allows identifying most relevant aspects for good or bad operational behavior.

2 Overview

We discuss the three steps stated in Section 1 to obtain the Reserving Excellence metric from a qualitative perspective (Section 3) as well as from a more detailed, quantitative perspective in (Section 4). Both the qualitative and the quantitative section discuss three parts related to determining the reserving difficulty of a claim, judging the reserving performance and combining individual aspects into a single metric. The qualitative part (Section 3) elaborates on basic assumptions and mechanisms in reserving, whereas the quantitative part (Section 4) states a model based on these assumptions and an assessment on actual data is done.

A reader who is unfamiliar with the area of reserving might resort to the definitions related to claims (see Section 4.2). These definitions are necessary for an in depth understanding of the quantitative part.

In Section 5 we provide a few examples to deepen the analytical understanding of the drill-down process and the metrics.

3 Qualitative Analysis

In this section we introduce essential characteristics on a qualitative level that provide the bigger picture for our later quantitative model and analysis.

3.1 Determining the Reserving Difficulty - Qualitatively

Here we answer the question: "What distinguishes a claim that can be easily reserved correctly from a claim that is difficult to reserve and requires frequent adjustments?" Our ultimate goal is to derive a schema that allows this question to be analyzed quantitatively such that we can assess the reserving difficulty for each claim individually. To this end, we start by presenting three key points resulting in several assumptions on a qualitative level:

- *Experience:* "How familiar is the claims handler (or company having a database of claims) with this type of claim?" We make the following assumption:

  Assumption 1 *The more experience a claims handler (or the company as a whole) has in handling a specific (kind of) claim, the less difficult a claim is to reserve.*

  The underlying reasoning is that there are correlations between claims that can be observed in our data that help in setting reserves. A claims handler might recognize certain patterns based on priorly handled claims. IT systems allow searching for similar claims that have already been dealt with. To see the practical impact of experience let us give an example. Motor accidents involving only damaged cars are very common. In contrast, an event such as the terror act of 9/11 is unique. Many insurance companies have standardized the process for reporting and computing reserves on motor accidents, since they are so common. For issuing a standard claim, there might be an online platform, where an insured is asked for all relevant information regarding his or her loss. The system might a database with repair costs for certain car parts of certain brands. Therefore, computing the reserves is often straight forward, fast and accurate.

  On the contrary, when opening a new line of business, no historical information is available in IT systems and also reserving actuaries might not be familiar with the particularities of this line of business. Therefore, it is natural that the company with its employees undergoes a learning phase, which might be characterized by more errors and lower productivity.

- *Predictability:* “How well can the claim be reserved given the information available about the claim, insured etc. as well as information about external factors?” It is generally more difficult to predict the more distant future than the close future. Thus, claims that are open for a long time are generally more difficult to predict than claims that are closed very shortly after their opening.

  Assumption 2 *Claims with a long life time are harder to reserve than short lived claims.*

  An insurance company has only limited information regarding a claim, its insured and external factors influencing the claim. The more unknowns there are, the more variability is inevitably also present in the reserve estimate. Often, the reason why claims remain open is that they cannot be settled due to unknowns or missing information which are hard to obtain or predict. For example, the outcome of litigation is uncertain or doctors might not be able to assess the long term impacts of an injury that is subject to a liability claim. An insured might not disclose some information intentionally to benefit from the inherent information asymmetry between insurer and insured. He might do at a later point in time when confronted in court or when investigations reveal the hidden information. A person buying life insurance might not disclose all information regarding his or her medical history to lower the premium. A person suffering disability due to an accident covered by workers compensation might receive monthly payments by the insurance until the end of his life. However, this person might die significantly earlier than expected or advances in technology and medicine might contribute to reduce its disability. All these examples illustrate that there is inherent uncertainty in reserving and it is generally impossible to reserve every claim correctly. In particular, claims that are open over years or even decades.

- *Complexity:* “How intricate is the claim?”

  Assumption 3 *The more complex a claim is the more difficult it is to reserve.*

  Losses involving multiple parties, unclear regulations regarding coverage etc. can be very difficult to deal with. A car accident involving several cars with many severely insured people and contradicting statements regarding the details of course of events might require much more investigation than a small damage at a car caused while parking. It seems evident that a complex claim is more difficult to reserve.

3.2 Judging Reserving Performance - Qualitatively

We describe several aspects that impact the reserving performance. They can also be used as individual metrics. The most obvious criterion to judge reserving performance is the area between the curve of the actual reserves and the curve giving the perfect reserves (shown as red area in Figure 3). Rather than simply taking this area we provide a more fine-grained analysis.

3.2.1 Duration and Magnitude of an Error

Assumption 4 *A small error over a long period of time is better than extreme changes in the incurred over a short period of time.*

For illustration, consider Figure 4 or the following example: The incurred of a claim A is set to 99% of the perfect reserves of claim A and it is never adjusted up to just before the last day before closure. The incurred of a claim B is 100% correct but for a relatively short period of time the incurred is set to just 20% of the perfect reserves of B. Out of the two claims, claim A has better Reserving Excellence. The reason is that it is almost impossible to exactly estimate the final incurred of a claim, whereas large deviations can be more easily avoided by proper analysis. More concretely, assume that an insurance company has to pay the repair of a car. The claims handler reserves 1000 USD on the registration date of the claim. The claim is closed on the day the final bill is received which amounts to 990 USD. This is preferable to keeping the claim reserves at 990 USD except for a short amount of time for which it is adjusted to 500 USD.

3.2.2 Time to Closure

It is evident that there is a causality between the information available about a claim and the ability to predict it. The more information at hand the less difficult a claim is to reserve. Thus, one would assume that errors shortly before closing should be treated as more severe, since generally more information is available than upon opening the claim.

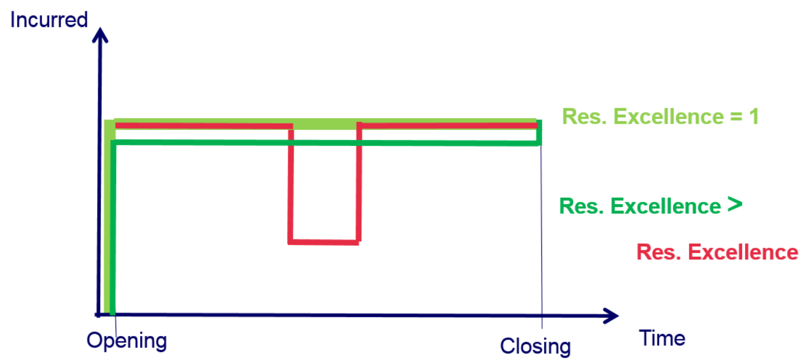

Figure 4: Since it is almost impossible to exactly estimate the final incurred of a claim, the Reserving Excellence is larger for a claim with a small error over a long period of time than for a claim with a large error for a short period of time.

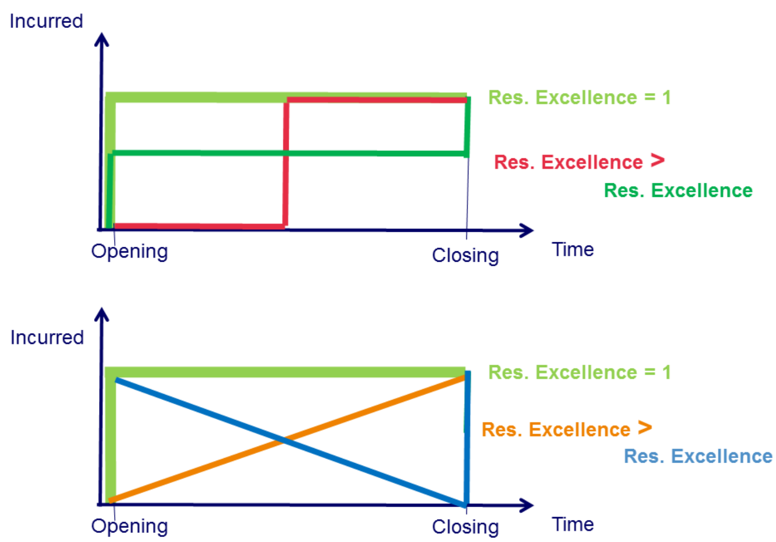

Figure 5: The reserving should be more accurate for claims near their closure dates. Thus, the Reserving Excellence for a claim A does not just depend on the area between the ideal setting of incurred and the (actual) incurred of claim A.

Assumption 5 *Errors in reserving shortly before closing a claim are worse than errors early in the life of a claim.*

For example, it should be easier to set the total incurred correctly for a claim the last day before its closing than two years before its closing. This implies that the Reserving Excellence of a single claim is not directly proportional to the area between its incurred over time and the ideal incurred, but it is weighed by the time to closing. For illustration, consult Figure 5. Note, that this statement is very different from saying that a claim with long lifetime is more difficult to reserve than a claim of short lifetime. Here, we say that for a single claim, it is more "tolerable" to make reserving mistakes shortly after its opening than shortly before its closure.

3.2.3 Number of Adjustments

With regard to accuracy of reserves, the reserves should be updated as often as possible. However, the total work related to adjustments is less if a single big adjustment is made rather than several small adjustments. This is motivated by the fact that each adjustment comes with a certain overhead, eg. a claims handler must look at a claim file and familiarize with it before being able to decide upon the amount of reserve adjustment. Thus, there is a trade-off between accuracy of results and frequency of adjustments. For an example, see Figure 6. Therefore, adjustments due to new information might be delayed in the hope that soon even more information is available. This might be desirable to minimize the amount of work done by a claims handler, but it is not desirable from the point of view that the current estimate of the incurred should reflect all available information about the claim.

For that reason if the focus of a company is on optimally estimating the incurred of claims, one might disregard the number of adjustments. However, if the focus is more in terms of efficiency in claims handling, the number of adjustments might be taken into considerations. There might also be company wide strategic goals, eg. to minimize reworking meaning "Get things right the first time." that might foster the inclusion of this metric. Since the choice is not clear per se, we do not include the number of reserve adjustments in our model.

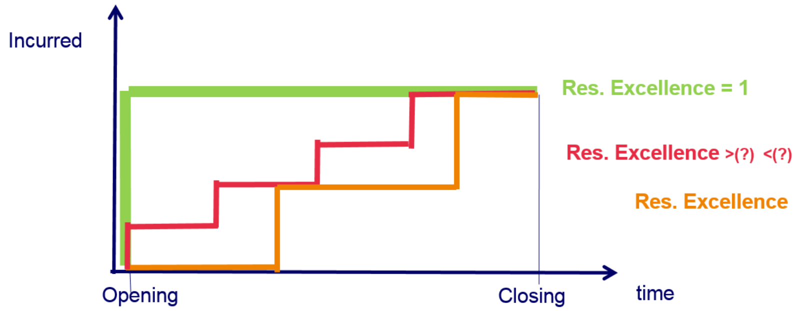

Figure 6: Frequent adjustments generally result in more precise reserving but more work for the claims handler. Which option is better depends on the preferences of a company.

3.3 Reserving Excellence - Qualitatively

For a claim we first compute its reserving difficulty followed by its reserving performance. The combination of both yields the Reserving Excellence. The reserving difficulty of a claim is used for weighing the reserving performance, ie. we value handling a difficult well more than handling a simple claim well. A key question is whether one should also include the importance of a claim. Consider two claims that are equivalently difficult to handle but one claim deals with a large loss and the other one only with a small loss. From a financial point of view a large loss is clearly more important. One might also argue that the (relative) error of two claims adjusted for complexity, experience and life time should be the same. From an operational point of view, a fifty per-cent error in reserves for a standard, simple claims of a small loss appears worse than a 10% error in a complex, long lived claim of large incurred. Therefore, we do not consider the importance (or absolute errors in reserving). Clearly, this assumption can be challenged and our model could be altered to capture such characteristics.

Assumption 6 *Each claim contributes to the operational excellence with a limited value.*

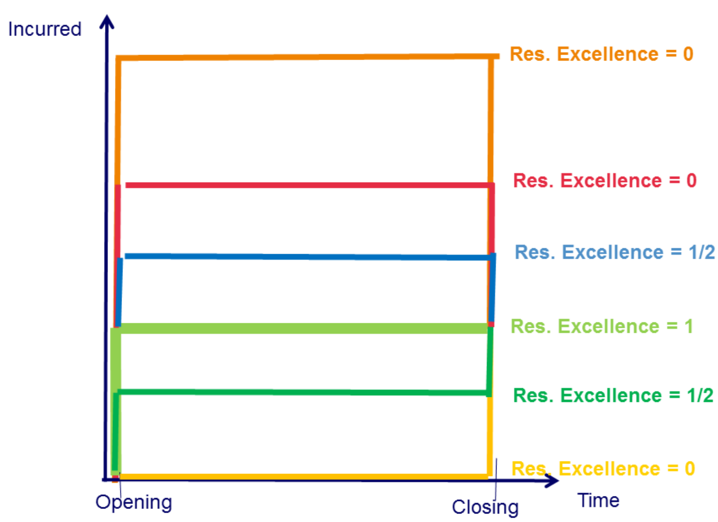

Figure 7: Reserving Excellence for extreme cases of incurred. The value is set to be between 0 and 1. All claims that are over-reserved by more than 100% of the ideal get the worst rating, ie. 0.

The assumption states that we assign each closed claim a value between zero and one for its Reserving Excellence, where one indicates a perfectly handled claim and zero indicates that the estimates of the incurred were done extremely badly.[2] This is shown in Figure 7. For an under-reserved claim a value of one corresponds to a claim with final incurred larger than zero but estimated incurred of zero throughout its lifetime. For an over-reserved a value of one indicates that (on average) the estimated incurred was double the actual incurred. Since the Reserving Excellence cannot go below zero this means that even if the incurred was larger, eg. triple or more of the incurred on close, then Reserving Excellence would still be zero. The underlying reasoning is that in case of extreme events that make a claim unpredictable the metric should not be distorted too much and, furthermore, a single drastic mistake for one claim should not influence the metric for all claims too much. However, this assumption could also be challenged.

4 Model and Quantitative Analysis

In this chapter we describe our model. From a mathematical point of view, we rely on standard linear regression to capture individual aspects contributing to the overall reserving performance of a claim (Section 4.3) and the reserving difficulty of a claim (Section 4.4). Linear regression is well established in economics and due to the linear dependence on individual factors the model provides relatively easy to grasp intuition how individual explanatory variables influence the overall metric. Combination of these factors into a single metric using weighing functions is conducted using different domain specific formulas that only rely on simple operations like addition and division. They are inspired by qualitative observations and priorly claimed assumptions. Thus, to understand the overall metric, it is essential to understand all qualitative aspects (Section 4) as well as details of the reserving process and its definitions (Section 4.2).

4.1 Data, Confidentiality and Model Validation

We used a large amount of claims data from different countries covering several years. The financial data included all relevant fields for each claim of a country mentioned in Section 4.2. Due to corporate policy all data in this paper has been distorted, such that only qualitative statements can be made. In particular, we refrain from giving details for model validation. Furthermore, no information regarding business units, line of business, country etc. can be given. As becomes evident from the discussion throughout the paper, the reserving process is complex and influenced by many external factors. Therefore, any metric must be interpreted with great care.

4.2 Definitions

A *claim* is defined as demand by an insured for indemnity under an insurance contract. A claim has a *Date of Loss*. The actual date of the loss depends on the trigger of the policy, eg. date of the cause for the claim or date of the insured incident. A claim also has a *notification date*, ie. the date the claim is reported to an insurer as well as a *registration date*, ie. the date the claim is recorded into the local claim system. A claim is considered *open* (sometimes denoted as outstanding), once it is registered, ie. the registration date and the *opening date* t^C_{open} coincide. We say that the *initial reserves* for the indemnity must be set, once a claim is registered. While a claim is open, the reserves might be altered. For simplicity and without loss of generality, we say that the (gross) *incurred* $I(t)^C$ of a claim at time t equals the reserves $R(t)^C$ plus payments $P(t)^C$ at time t. Thus, we do not further discuss the impact of deductibles, recoveries, reinsurance coverages etc. Therefore, while a claim is open its *incurred* might change due to changes in reserves or due to payments. However, for a perfectly reserved claim, the incurred does not change, ie. an increase in payments goes hand in hand with a decrease by the same amount in reserves (see Figure 2). An open claim is eventually settled and *closed*. For a closed claim, the reserves are zero, all payments have been made and, consequently, the incurred $I(t^C_{close})$ remains unchanged. The incurred before the closing date is an estimate of the total payments $P(t^C_{close})$ that have to be made. The date on which a claim C is changed from open to closed is called *closing date* t^C_{close}. The timespan between the opening and closing date, ie. $t^C_{close} - t^C_{open}$ is called *life time*. At time t_0 a claim is of *age* $t_0 - t^C_{open}$. A claim is *perfectly reserved* if the incurred of the open claim remains unchanged and equals the incurred of the closed claim. For illustration of the behavior of incurred, reserves and incurred as well as for a perfectly reserved claim see Figure 1. Since the incurred equals payments plus reserves, this implies that after the initial reserves are set, (ideally) reserves are only changed due to payments. More precisely, whenever a payment is made of an amount X, the reserves are reduced by the same amount X. A

[2] However, one might also argue that a complex claim results in more work and therefore, should get more weight. The amount of work might for instance be estimated using the handling costs of a claim.

Figure 8: Average percentage of changes in incurred relative to final incurred versus the age of a claim for claim. All claims have the same lifetime.

closed claim might be *reopened*. For simplicity and to focus on key ideas, we do not discuss reopened claims here, but assume that a closed claim remains closed.

A time interval $[t_0, t_1]$ is given by all times t with $t_0 \leq t \leq t_1$. Let all claims that have ever been registered be S. All closed claims within interval $[t_0, t_1]$ are given by $C_{closed}(t_0, t_1) := \{C \in S | t_0 \leq t^C_{close} \leq t_1\}$. The number of all closed claims within interval $[t_0, t_1]$ is given by $|C_{closed}(t_0, t_1)|$.

We define the change of incurred in percent for a claim C as $ChangeInc(C) := \sum_{t^C_{open}}^{t^C_{close}} |I(t)^C - I(t-1)^C| / I(t^C_{close})$. The average change of incurred in percent for a set of claims S is then $\sum_{C \in S} ChangeInc(C)/|S|$.

4.3 Judging Reserving Performance - Quantitatively

For a particular age t of a claim we compute one term capturing the deviation of incurred and one term that accounting for (relative) age of a claim.
To compute deviation of incurred we use the L2-norm to account for the fact that small errors over a long time are more tolerable than large errors over a small time as shown in Figure 4 and based on Assumption 4. Motivated by Assumption 6 we use a normalized (rather than an absolute) gap between the incurred $I(t)$ at time t and the final incurred at closing time, ie. $\frac{(I(t^C_{close}) - I(t))^2}{I(t^C_{close})^2}$. Since the final incurred $I(t^C_{close})$ might be zero[3] we must limit the factor. Furthermore, if both incurred $I(t^C_{close})$ and $I(t)$ are zero, the ratio is undefined. To deal with this, we might add a small number (eg. 0.000001) to the final incurred such that it cannot be 0. However, we follow a different approach, ie. we use

$$f_1(t) = \begin{cases} \frac{(I(t^C_{close}) - I(t))^2}{I(t^C_{close})^2} & \text{if } I(t^C_{close}) > 0 \\ 0 & \text{if } I(t^C_{close}) = 0 = I(t) \\ 1 & \text{if } I(t^C_{close}) = 0, I(t) > 0 \end{cases}$$

This also limits the Reserving Excellence as illustrated in Figure 7.

To get a better understanding of the impact of time to closure (Section 3.2.2) consider Figure 8. The number of adjustments as well as the severity of the adjustments should decrease over time according to our hypotheses, since more information is available about the claim. Our empirical analysis (see Figure 8) confirms that severity of adjustments decreases the older a claim gets. Note that just before closure there are frequently big adjustments being made. For example, a claim might be closed because the result of litigation or the bills of repair work are available.

We fitted a linear function $f_2(t) := k_0 \cdot t/(t^C_{close} - t^C_{open}) + k_2$ using regression obtaining the constants k_i. The weighing by age t accounts for the fact that initial errors are preferable to late errors (Assumption 5). Note that

[3]Any claim creates administrative costs. However, from an accounting perspective, these costs might be unallocated loss adjustment expenses (ULAE), rather than costs that are directly attributed to a specific claim.

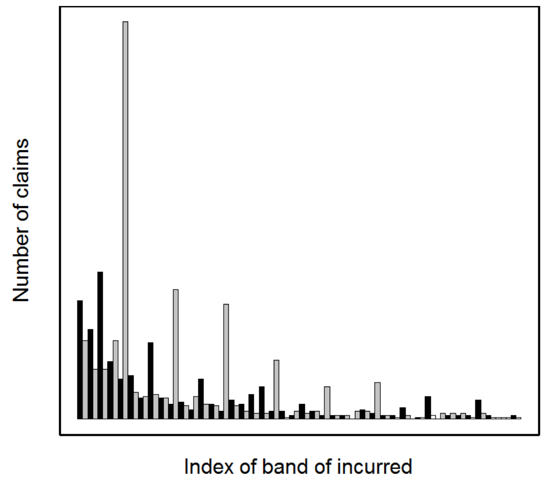

Figure 9: Number of claims vs bands of incurred (of equal width) ignoring 5% of claims with largest incurred of an exemplary country.

we normalize by dividing the age by life time. The assumption 2 stating that long lived claims are more difficult to handle is included in the reserving difficulty. The overall reserving performance is defined to be the product of $f_1(t)$ and $f_2(t)$, ie. $f_1(t) \cdot f_2(t)$. Addition would have been an alternative, but we considered the scenario where there are rather larger errors in reserving towards the end of a claim as very severe. This severity is better expressed using multiplication.

4.4 Determining and Modeling the Reserving Difficulty - Quantitatively

We are interested in finding criteria to assess the reserving difficulty of a claim based only on the financial reserving history as well as on the claim counts using our previously mentioned dimensions experience, predictability and complexity. Our underlying hope is that claims that are similar to claims with respect to their available financial information (ie. incurred) are also similar with respect to their reserving difficulty (for a particular line of business). Given this holds the key problem is finding a means to determine the similarity upon claims using the available data. Here, we use financial information and claim counts. This data is an outcome of business practices and accounting principles that make the data dependent on the company, line of business and country.

We relate qualitative measures and assumptions (Section 3.1) to quantitative measures of the data and provide a brief assessment of the relationship using actual data:

1. *Experience:* Claim counts relate to experience. The more claims have been handled the more experienced a claim handler (or company as a whole) becomes. We use claim counts for bands of incurred. As can be seen in Figure 9 large losses are less frequent than small attritional claims. Therefore, the company as a whole has more experience in claims consisting of small losses. Note that the incurred does not follow a smooth distribution. There are several reasons for this behavior. For example, the deductibles might alter the distribution (depending on how the incurred is computed exactly). There are peaks at certain levels. Peaks have different reasons. For instance, people tend to round numbers. Claims might be closed without payments to cover the loss. But payments might be due to costs related to other factors, eg. due to administration costs or costs due to investigation upon the claim, eg. to check the validity of a claim before denying payment. The distribution shown in Figure 9 is probably best described by a Weibull or an exponential distribution. However, as we shall discuss later, we ignore claims with small incurred and use a linear function for the remaining claims.

2. *Complexity:* Incurred (generally) relates to complexity. For example, large losses are often complex and they are frequently handled by special teams. The assumption is that due to the increased complexity we expect relatively larger adjustments for claims with large incurred. Actual data only partially confirms this assumption (see Figure 10). Claims with small final incurred are adjusted drastically (in percentage of their final incurred). Then there is a small level of adjustment followed by a trend showing that the percentage of adjustments increases with the final incurred. A key question is: "Why are adjustments for claims with small incurred large?" Due to their frequency of occurrence (Figure 9) one would expect standardized processes and given our assumption of an increase of complexity with incurred they should show the least degree of

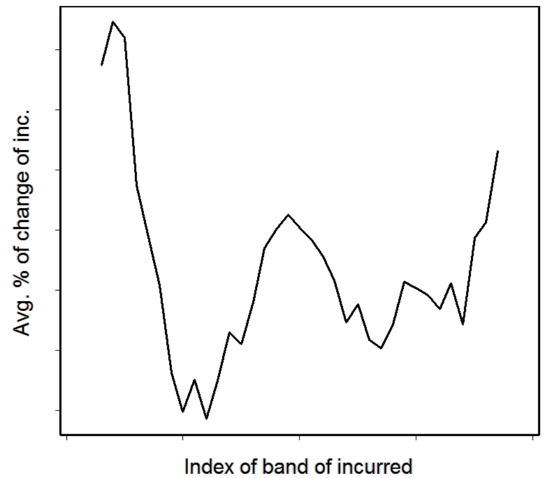

Figure 10: Average change of incurred in percent per band of incurred of a country.

adjustments. There are a variety of possible explanations. Note that people have less incentive to report very small accidents and since many insurance policies have deductibles claims with small losses might be reported less. Nevertheless, there are claims (Figure 9) with small incurred. Many claims are due to the fact that payments to the insured are declined (which happens rather frequently) but still administrative or investigative costs apply. Depending on the country and line of business they might significantly exceed 10% of all claims. Claims closed without payment to the insured have a strong impact. These claims might originally be reserved conservatively while there was still a reasonable degree of uncertainty with respect to their final incurred, followed by a large change in reserves due to additional insights or legislative decisions. There could also be reasons due to internal claims handling processes. A claims handler might first do a rough estimate of the incurred of a claim. Based on a threshold claims with small estimates are not treated further until more information is available, i.e. a repair bill of a car is being issued. The process might demand that claims above the threshold are treated with more effort resulting in better estimates. Therefore, since the estimate of small incurred is rather close to a guess, it is expected that these claims must be adjusted more (relative to their final incurred) than claims with larger incurred. Clearly, such a process targets to minimize costs, since putting effort into adjusting claims of small incurred might not be worth-while. For simplicity we ignore claims of small incurred. When doing so, it can be seen that there is indeed a positive correlation of the incurred and the amount of change of a claim (see Figure 10).

3. *Predictability:* Predictability cannot be inferred well from financial information only. However, predictability correlates with the lifespan of a claim. Predictability is harder for claims with a long life time. In other words, claims that are open for a long time before closure are generally more difficult to reserve than claims that are settled shortly after reporting as stated in Assumption 5. Therefore, if two claims are reported and adjusted at the same points in time but the first claim is closed much after the second then the first claim with longer lifespan has better Reserving Excellence.

   The relation between the lifetime of a claim and its percentage wise adjustment relative to the final incurred is shown in Figure 11. We also see a trend that the adjustments increase with lifetime. Claims that have been open for a long time are adjusted more than claims that are handled (ie. closed) quickly. However, the trend is not as smooth as one might expect. There are many factors that can contribute to such irregularities. For example, legislation procedures might often take between 8 to 16 months in a certain country for certain types of cases relevant to the line of business shown. Such cases might be very difficult to predict. Still overall we see a positive correlation of life-time of a claim and number of adjustments.

Next, we derive a model for the reserving difficulty to unite these findings. To obtain the reserving difficulty, we must also account for correlation between the incurred and life time of a claim. For simplicity, we use a linear model to fit the constants k_i using standard linear regression of the following model:

$$w(Inc, LifeTime) := k_0 \cdot Inc + k_1 \cdot LifeTime + k_2 \cdot Inc \cdot LifeTime + k_3$$

This model captures the trends shown in Figure s10 and 11 (ignoring claims with small incurred). Though, using statistical software it is not hard to fit more complex models, to develop a basic understanding a linear model is most suitable and it also reduces the risk of overfitting. Furthermore, we limited the reserving difficulty to one

Figure 11: Average percentage of changes in incurred relative to final incurred versus the life time.

in order to avoid that outliers distort the metric too much (in accordance with Assumption 6). This gives us the reserving difficulty of a single claim. One might also consider the number of adjustments rather or in addition to the severity of adjustments.

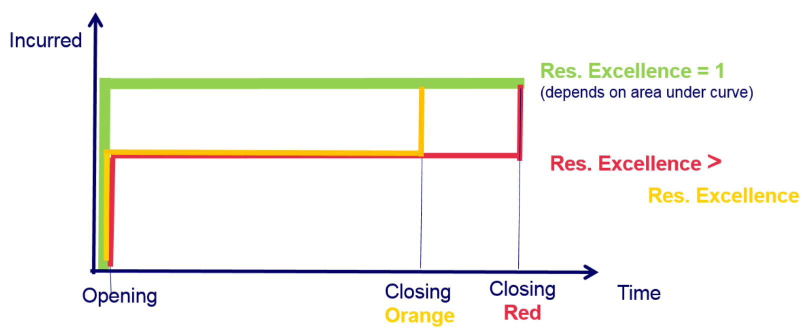

Figure 12: Given all things equal, the Reserving Excellence is better for claims with larger life time.

4.5 Determining the Reserving Excellence - Quantitatively

Let us consider the Reserving Excellence of a single claim C as displayed in Figure 3. Let us compute the contribution of time t to the entire Reserving Excellence. The reserving performance is given by the product of functions $f_1(t)$ and $f_2(t)$ (see Section 4.3). Including the judgment $w(Inc, LifeTime)$ of the difficulty of a claim (see Section 4.3), the Reserving Excellence for a closed claim C is defined as:

$$ResExc(C) := 1 - (1 - w(Inc, LifeTime)) \cdot \min(1, \sum_{t=t^C_{open}}^{t^C_{close}} f_1(t) \cdot f_2(t))$$

We define the Reserving Excellence over time, ie. evaluated at times t_i as follows. Let S_i be all claims that have been closed between $]t_{i-1}, t_i]$, then the Reserving Excellence is just the average of all claims S_i:

$$ResExc(t_i) := 1/|S_i| \cdot \sum_{C \in S_i} ResExc(C)$$

5 Case Studies and Illustrations

We illustrate the Reserving Excellence metric for the model described in Section 4. The provided examples touch upon all aspects regarding the reserving difficulty as well as determining the reserving performance. We begin with single claims on a conceptual level and then move to aggregates. Finally, we also discuss aspects related to actual data.

5.1 Single Claim

We discuss several scenarios relating mainly to the interpretation of the Reserving Excellence of a single claim.

5.1.1 Judging the Reserving Difficulty

Let us give some intuition regarding the model in Section 4.4. In Figure 13 the impact of different life times of claims is illustrated. There is a linear behavior between the life time of a claim and the estimated reserving difficulty. The relationship in Figure 14 is more complex, since two considerations come into play. First, for larger incurred the reserving difficulty is increased, since claims with large incurred are harder to reserve correctly. Second, the larger the relative difference between the estimated incurred and the actual incurred, the smaller the Reserving Excellence. Both aspects are most profound for Claim A: The final incurred is small classifying the claim as easy to reserve(non-complex), the relative error in the estimate is large showcasing bad reserving (in fact, it is unbounded, since it jumps from zero to ten). For claim E the relative reserving error is small (compared to A) and the incurred is rather large, resulting in the claim to be judged as well-reserved, ie. the Reserving Excellence is close to 1.

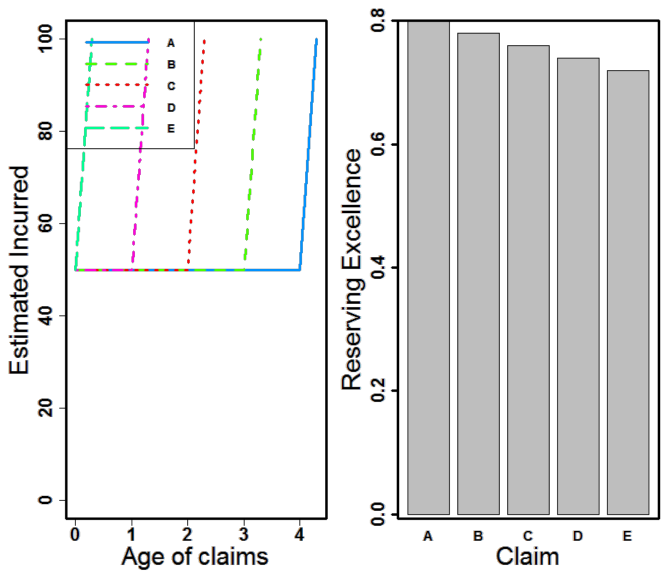

Figure 13: Reserving Excellence for claims with different life time but same constant offset of incurred from the perfectly reserved claim.

5.1.2 Judging the Reserving Excellence

Figure 15 shows claims that have a constant error. The reserving difficulty is set to be the same for all shown claims. The linear relationship between the Reserving Excellence and constant offset of the claims can easily be seen. Whether a claim with an offset of 100% should get a Reserving Excellence of 0 as in Figure 15 is a matter of taste. One could also fix the scale, such that (the best) 5% of all claims get Reserving Excellence equal to one and the worst 5% get a zero.

Figure 16 exhibits non-linear behavior in the Reserving Excellence due to a linear change of the incurred over time. Claim A has received a larger Reserving Excellence than claim B since it converges towards the correct incurred, whereas B diverges (see also Section 3.2.2). The same reasoning applies for claims C and D. It is interesting to compare claims A, B, C, D with claim E, which has a constant offset. Here, two considerations come into play. First, large deviations from the optimal diminish the incurred more than proportional due to usage of the L2-norm (see Section 3.2.1 for motivation). Second, deviations from the optimal incurred reduce the Reserving Excellence of a claim more if they occur near the closing of a claim. For claim B the first aspect is larger than the second, thus it is rated worse than claim E. However, for claim C the second aspect outweighs the first and it is rated slightly (non-visible in figure 16) better than claim E in terms of the Reserving Excellence.

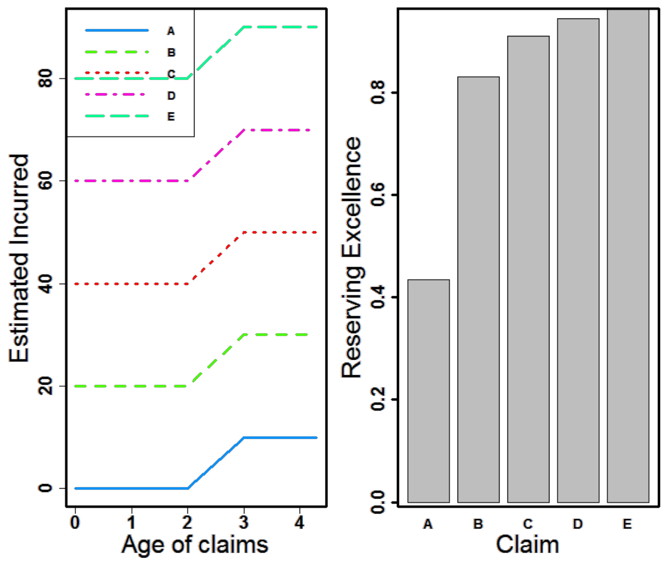

Figure 14: Reserving Excellence for claims with vertically shifted incurred curves.

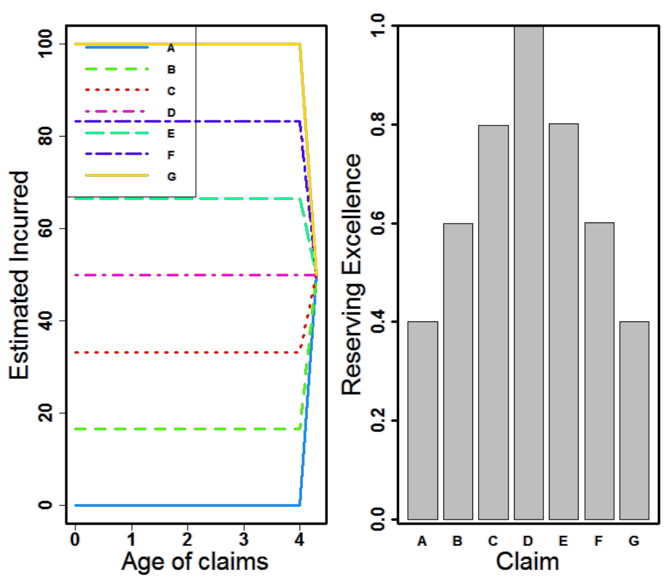

Figure 15: Reserving Excellence for claims with constant offset of incurred from the perfectly reserved claim but equal reserving difficulty. The (correct) incurred is 50 for all claims.

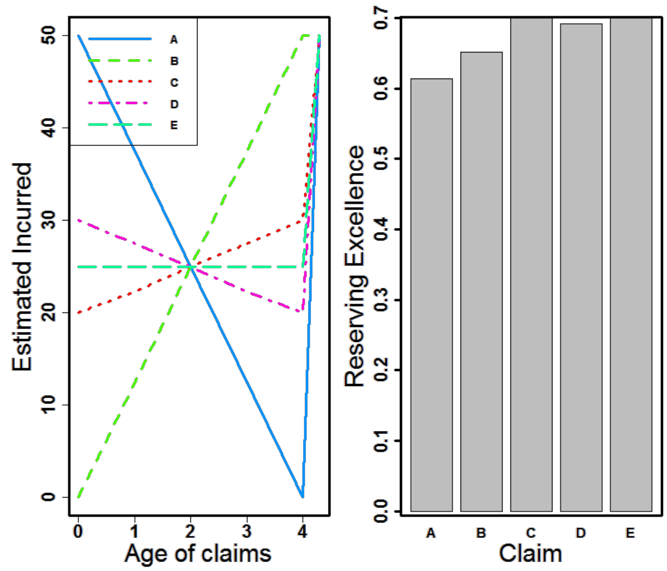

Figure 16: Reserving Excellence for claims with linear offset of incurred from the perfectly reserved claim. The value is set to be between 0 and 1. The final incurred is 50 for all claims.

5.2 Aggregates of Several Claims

We consider the Reserving Excellence of several claims over time. The goal is to look at several cases for which we perform a drill-down and determine the factor which caused the change of the Reserving Excellence, ie. one or several factors related to reserving difficulty or performance judgment. In other words, we do not just compute the average Reserving Excellence over time, but also the underlying factor, such as average reserving difficulty and average performance judgment. One could consider more factors, eg. drill-down into the performance judgment by looking at the averages of $f_1(t)$ and $f_2(t)$. Using traditional metrics such as the average lifetime of a claim, the average incurred of a claim, the number of closed claims might yield insights as well. Furthermore, looking at these metrics per band of incurred could also be beneficial. In the examples, we evaluate the Reserving Excellence only at two points in time and use just two claims for each evaluation point. More precisely, at time t_0 we consider all claims that have been closed at t_0. Generally, the state of claims is updated and stored with fixed intervals, eg. on a monthly basis. Thus, for an adjustment we might not know the exact date but only that it occurs in a particular month. In all following figures we write claim X closed $@t = t_0$ to denote a claim that has been closed in some interval up to t_0, eg. $[t_0 - 1(month), t_0]$.

Figure 17 shows the impact of claims having longer life times. Each claim has the same difference between the estimated incurred and the actual incurred up to the time it is closed. Claims A, B were closed at time $t = 1$ and claims C, D at time $t = 2$. Thus, claims A, B are used to compute the Reserving Excellence at time $t = 1$ and the claims C, D for time $t = 2$. However, claims C, D have longer life times than claims A, B. Thus, the Reserving Excellence increases due to an increase in reserving difficulty, whereas the reserving performance stays constant.

Figure 18 shows the impact of claims with reduced gap between actual incurred and estimate, ie. the claims C, D have a more (or equally) accurate estimate of the incurred than claims A, B at all times. However, claims C, D also have a smaller actual incurred than claims A, B. The latter implies a decrease in reserving difficulty, which would yield a reduction in the Reserving Excellence. But the reduction in the gap of the estimate and actual incurred improves upon the reserving performance, yielding an overall increase of the Reserving Excellence.

5.2.1 Actual Aggregates

Figure 19 shows the Reserving Excellence of a large number of claims over time. Before interpreting the model using the previously described techniques, it is important to be aware of irregularities in the data. There are fluctuations in the data that are inherent. More precisely, the number of claims in terms of counts and severity changes over time since the events causing the claims are stochastic in nature. This could have several impacts. For example, a larger amount of claims could mean more work for claims handler and less time for judging an individual claim, causing a larger error. Our current metric does not take into account the workload, ie. one might define Reserving Excellence also with respect to efficiency. Thus, one could include the number of claim handlers into the metric. Currently, we only measure Reserving Excellence based on the quality of reserving irrespective of the resources used

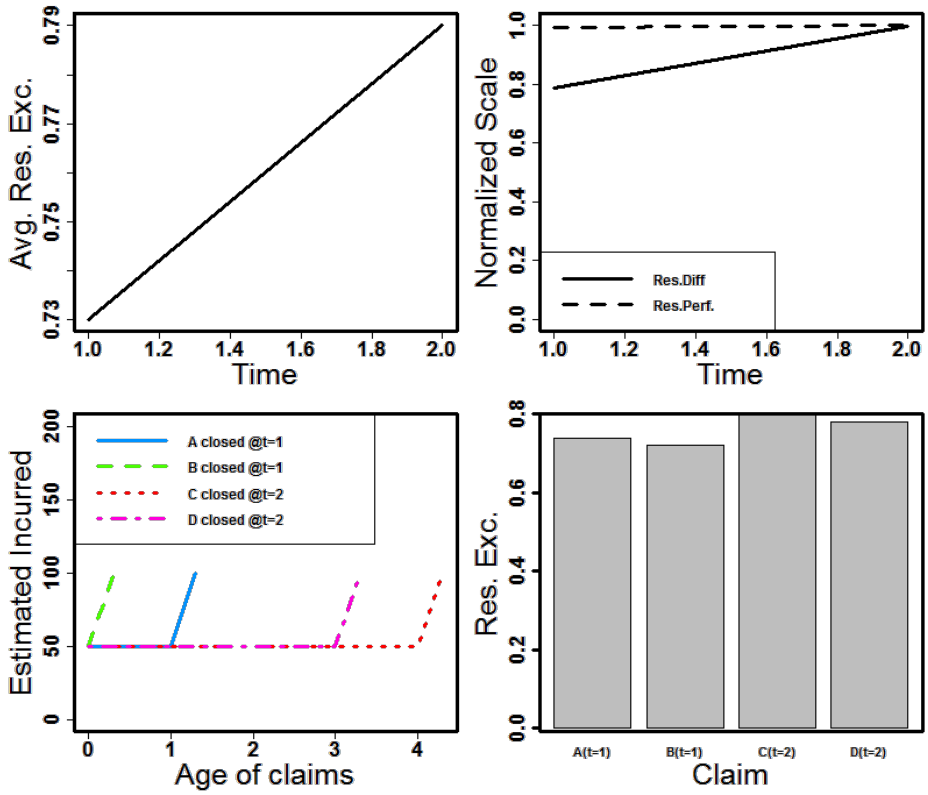

Figure 17: Average Reserving Excellence for claims with different life time

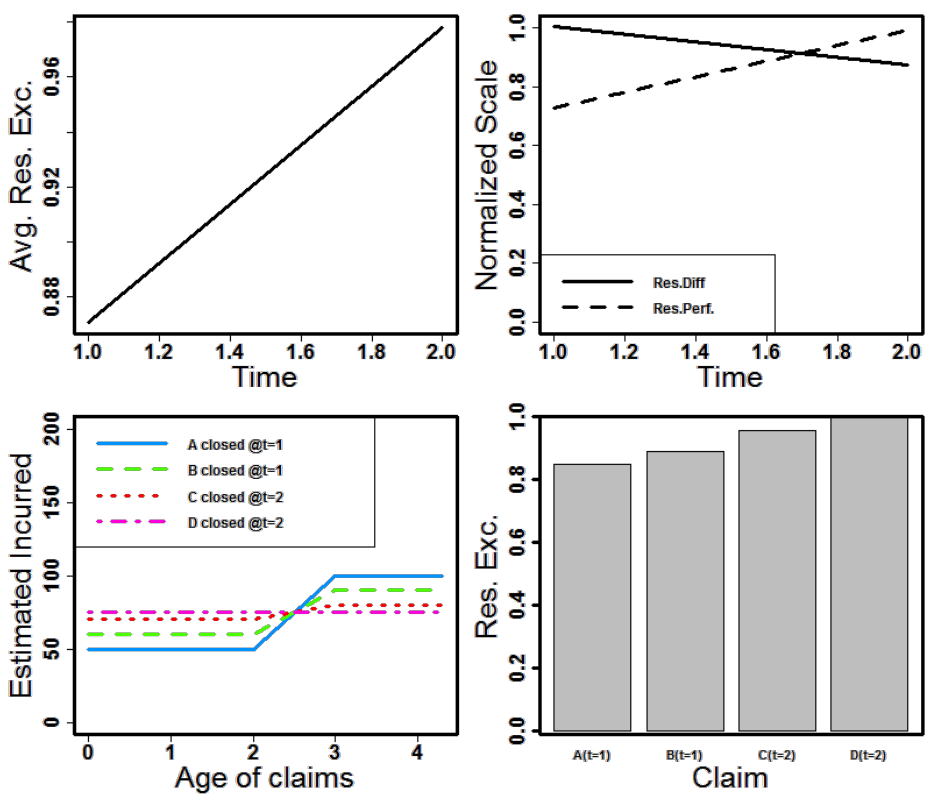

Figure 18: Average Reserving Excellence for claims with different gaps between estimated and actual incurred as well as different closing times

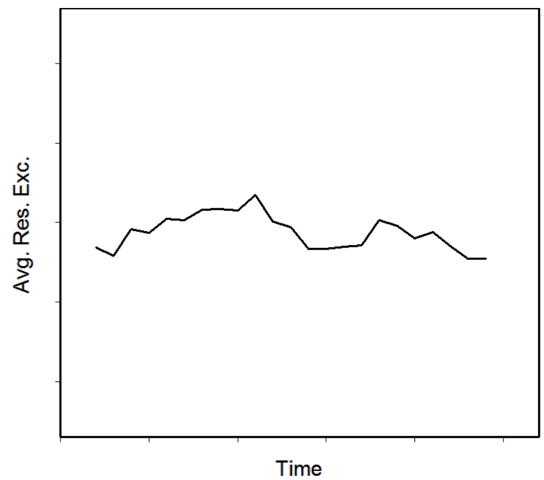

Figure 19: Reserving Excellence for actual data for several years and several thousands of claims

to conduct the reserving. Variation in severity should not have a strong impact on our model, since we weigh claims based on their severity. Furthermore, long term trends should still not be changed due to random fluctuations. However, there are also factors that alter the claims distribution that are related to business practices. For instance, the decision to use reinsurance for claims above a certain threshold of incurred or to adjust deductibles changes the distribution of the number of claims per incurred. Premiums might be adjusted to attract customers with a lower risk profile. A new competitor might enter the market attracting certain types of customers with a certain risk profile. Such decisions or events not only impact the outcomes of the Reserving Excellence, but potentially also the validity of the parameters and potentially the assumptions reflected in the model. Once confidence in the data has been established, the Reserving Excellence can be used to measure operational excellence over longer time periods but also to assess initiatives targeting the improvement of the reserving process.

6 Related Work

There exist a lot of methodologies that target to estimate the total incurred of claims and thus also the reserves correctly. For instance, the article [4] discusses stochastic claim reserving such as the Chain-ladder algorithm or Bornhuetter Ferguson algorithm based on aggregate data. Traditional (aggregate) models like Paid-incurred chain (PIC) claims are still subject to study, eg. [10] takes into account novel aspects such as solvency regulations. There is also work on modeling claims development processes for individual claims that make weaker assumptions regarding independence of claims than stochastic models, eg. [12, 8].

Reserves can be reported wrongly by intention, eg. to comply with solvency regulations or for smoothing purposes or tax avoidance [5, 7, 6]. Utility theory has been discussed in the context of reserving [13]. The authors describe interests and motivations of various parties, such as underwriting or claims, of setting the reserves too high, too low or correctly. Factors beyond lack of information or human error such as earnings can impact the accuracy of reserves. For instance, property-casualty insurers with small positive earnings understate loss reserves relative to insurance providers with small negative earnings [1]. In all of these articles [5, 7, 1, 6] reserve amounts are discussed on an aggregate, ie. company wide, level rather than on a single claims level, see related work of [6]. For example, [15] compares the originally reported loss reserve to (all) future claims paid. More precisely, the error Err_t of a firm at calendar year t is given by $Err_t := IncurredLosses_t - DevelopedLossesPaid_{t+j}$, where incurred losses are the total losses known to the insurer and, additionally, losses that have been estimated to have occurred. Developed Losses Paid at a future year $t + j$ are those losses actually paid by the insurer. [6] also introduces a reserve error metric based on stochastic loss reserving models. The paper [2] uses (primarily) the mean absolute percentage error statistic to compare the accuracy of several loss reserving methodologies. The motivation to stay away from a squared loss error is that for overall reserving accuracy (from a purely financial point of view), it does not matter whether a few claims are reserved with large errors or a few claims are off with a small error. This is in contrast to our work - see Section 3.2.1.

More generally, all the above presented loss error metrics are in stark contrast to our work, where errors in reserving are based on the errors of individual claims rather than a simple aggregate not weighing individual claims. Furthermore, aside from the financial deviation, we take into a account when an adjustment occurs in the life of a claim and we also consider the entire reserving history of a claim. Additionally, we do not account for IBNR claims, since we are only interested in how well existing claims have been reserved.

Despite the fact that certain key performance indicators are well-known in the insurance industry, the number of scientific publications is rather limited. Among these metrics are, for instance: Average handling cost per insurance claim, percentage of insurance claims handled first time correctly, Average number of insurance claims per handler, percentage of escalated insurance claims, Average closure duration of insurance claims, percentage of re-opened insurance claims, average incurred on reported claims and so on. The naming might differ from insurer to insurer. Often insurance companies set their own reporting standards including these or similar metrics with extensive documentation, eg. [14]. A less complete list of metrics can be found in textbooks, eg. [3], articles, eg. [9], or on publicly available presentations [11]. Most of these metrics are rather straight forward to calculate. The paper [9] discusses the impact and the actions to be taken by a reserving actuary to identify emerging effects from claim initiatives. Thus, the paper does not deal primarily with developing metrics but more with stating a methodology to deal with claim initiatives based on examples. It also states that (long term) reserving experience might be of limited value due to the impact of these initiatives. The work [2] compares several reserving methodologies based on accuracy (ie. using the mean absolute percentage error) and bias. It briefly discusses responsiveness and stability on a qualitative level. To the best of our knowledge, there are no metrics that explicitly deal with Reserving Excellence to the extent that we do, eg. taking into account the entire financial history of a claim and focusing on several aspects of claim reserves evaluation rather than simple averages.

7 Conclusion

We have stated a key performance indicator for operational excellence of reserving. It allows tracking the impact of initiatives targeting the reserving process as well as spotting problems in the reserving process by considering a single metric. Drill-down to individual claims is possible as well as looking at the composition of the KPI based on sub-metrics. However, some caution is required in the interpretation and in deriving the underlying model: Valid conclusions can only be made having adequate data in terms of quantity and quality. Furthermore, we explicitly excluded certain aspects of reserving that could also be included for computing the reserving excellence. Nevertheless, given the lack of metrics for operational excellence, we see our metric (and its derivation using sub-metrics) as a first valuable approach to allow for better quantitative management of reserving operations.

References

[1] W. H. Beaver, M. F. McNichols, and K. K. Nelson. Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. *Journal of Accounting and Economics*, 35(3):347–376, 2003.

[2] T. Boles and A. Staudt. On the accuracy of loss reserving methodology. In *Casualty Actuarial Society (CAS) Forum*, 2010.

[3] M. Denuit, X. Maréchal, S. Pitrebois, and J.-F. Walhin. *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons, 2007.

[4] P. D. England and R. J. Verrall. Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(03):443–518, 2002.

[5] J. J. Gaver and J. S. Paterson. Do insurers manipulate loss reserves to mask solvency problems? *Journal of Accounting and Economics*, 37(3):393–416, 2004.

[6] M. F. Grace and J. T. Leverty. Full information reserve errors and their relation to auditor and actuary quality. In *ARIA 2011 Anual Meeting*. Citeseer, 2011.

[7] M. F. Grace and J. T. Leverty. Property–liability insurer reserve error: Motive, manipulation, or mistake. *Journal of Risk and Insurance*, 79(2):351–380, 2012.

[8] J. Guszcza and J. Lommele. Loss reserving using claim-level data. In *Casualty Actuarial Society (CAS) Forum*, 2006.

[9] A. Halpert, S. Weinstein, and C. Gonwa. Evaluating reserves in a changing claims environment.

[10] S. Happ. Stochastic claims reserving under consideration of various different sources of information. *Dissertation, University Hamburg*, 2014.

[11] S. Investors day presentation, Zurich Insurance Company Ltd. Claims management and reserving. `http://www.zurich.com/internet/main/SiteCollectionDocuments/presentations/archive/20050630_investors_day_en.pdf`, 2005.

[12] C. R. Larsen. An individual claims reserving model. *Astin Bulletin*, 37(1):113, 2007.

[13] M. W. Littmann. Using utility theory for describing best estimate reserves.

[14] V. Salipante. Gi claims definitions reference guide. *Internal Report, 2nd Edition*, 2012.

[15] M. Weiss. A multivariate analysis of loss reserving estimates in property-liability insurers. *Journal of Risk and Insurance*, pages 199–221, 1985.