**Compiling Data** **20m**

Data can be categorized as coming from either **internal** or **external** sources. Large insurers often rely solely on data produced internally, as their own management information systems are capable of generating the detailed claims and exposure data required. In contrast, smaller insurers may face limitations in generating internal data, such as limited volume and system constraints, which force them to use external data. Large insurers that have recently expanded into new lines of business or geographical regions may also turn to external sources of data.

Regardless of whether the data is from internal or external sources, it is the actuary's responsibility to evaluate and select the data with due consideration to reasonableness, appropriateness, comprehensiveness, and other factors outlined in *Actuarial Standards of Practice No. 23, Data Quality (ASOP 23)*.

## Internal Data

Internal data can be divided into two categories:

1. **Risk information:** This includes information specific to a policy/claim, such as exposures, premium, claim counts, losses, and explanatory characteristics about the policy/claim.
2. **Accounting information:** This includes information that is not specific to any one policy, such as underwriting expenses and ULAE. Accounting information is tracked at the aggregate level.

### RISK INFORMATION

Risk information can be obtained from an insurer's policy database and claim database. A *policy database* contains information about policy exposure and premium. The definition of a record varies by the line of business. For example, a record for homeowners insurance may be one home insured for an annual policy period. For personal auto insurance, separate records may be created for each individual operator of the vehicle and each coverage.

Each record in the policy database usually includes the following fields:

- **Policy identifier:** This is a unique identifier that corresponds to a particular policy.
- **Risk identifier:** This is a unique identifier that corresponds to a particular risk, such as the vehicle number for personal auto insurance that insures multiple vehicles.
- **Relevant dates:** This includes the original effective and termination dates for the policy or each risk/coverage and any dates of amendment.
- **Premium:** This is typically the written premium associated with the record.
- **Exposure:** This is typically the written exposure associated with the record.
- **Characteristics:** These fields can include rating variables, underwriting variables, and any other available information regarding the risk represented by the record.

A ***claims database*** captures claims information on a specific policy. Typically, each record within the claims database corresponds to a transaction associated with a specific claim.

Each record in the claims database usually includes the following fields:

- **Policy identifier:** This is a unique identifier that corresponds to a particular policy.
- **Risk identifier:** This is a unique identifier that corresponds to a particular risk.
- **Claim identifier:** This is a unique identifier that corresponds to a particular claim.
- **Claimant identifier:** This is a unique identifier that corresponds to each specific claimant on a particular claim.
- **Relevant loss dates:** Including the date of loss, the report date, and the date of a transaction (e.g., date of a loss payment, reserve change, or claim status change).
- **Claim status:** This field indicates whether the claim is open (active), closed (settled), re-opened, or re-closed.
- **Claim count:** This field identifies the number of claims by coverage associated with the loss occurrence.
- **Paid loss:** This is the amount paid for a claim transaction. When there are multiple coverages, perils, or types of loss, the loss payments can be tracked in distinct fields or separate records.
- **Event identifier:** This indicates if the claim involves any extraordinary event, e.g., a catastrophe.

- **Case reserve:** This is either the new case reserve or the change in the case reserve.
- **Allocated loss adjustment expense:** The expenses that can be directly assigned to the specific claim.
- **Salvage/subrogation:** This is the amount of recovery from salvage or subrogation.
- **Claim characteristics:** Any other information about the claim that could be relevant to the actuary, such as the type of injury.

## External Data

External data is often used to price a new line of business or as a supplement to internal data when pricing an existing line of business. External data can also be particularly valuable when selecting tail development factors, trend rates, and expected claim ratios.

The most commonly used sources of external information include data calls or statistical plan data, other aggregated insurance industry data, competitors' rate filings, and third-party data unrelated to insurance.

### STATISTICAL PLANS

In the United States, P&C insurance companies are often mandated by regulators to file statistical data in a consistent format. The required statistical plan is typically a summary-based plan, so the data does not need to be detailed.

An example of a statistical plan is ***The Texas Private Passenger Automobile Statistical Plan***, which is established by the Texas Commissioner of Insurance. Companies are required to submit data that is aggregated by territory, deductible, and driver class. This statistical plan is used as a basis for setting benchmark rates for personal automobile insurance premiums and as a supplement for companies performing internal analyses on personal automobile insurance.

Besides state regulators, some industry service organizations also collect and aggregate data from a group of participating companies writing the same insurance product. The **National Council for Compensation Insurance (NCCI)** and **Insurance Services Office, Inc (ISO)** are examples of such organizations. Other than collecting and summarizing data, these organizations also analyze the aggregated data and provide the results of their analysis to the participating companies. Unlike state regulators, industry service organizations collect data at the transactional level so that they can perform detailed actuarial analyses of the data.

## OTHER AGGREGATED INDUSTRY DATA

Other organizations that collect, aggregate, and analyze insurance data include:

1. Fast Track Monitoring System
    - Collects and analyzes loss data on personal lines of business in the United States
2. Highway Loss Data Institute (HLDI)
    - Sponsored by a group of U.S. personal auto insurers and insurance associations
    - Compiles insurance data reported by member companies
    - Provides detailed loss information categorized by type of car to both member companies and public policymakers
    - Offers summarized information that can be beneficial for insurers and the general public
3. Insurance Research Council (IRC)
4. Institute for Business and Home Safety (IBHS)
5. National Insurance Crime Bureau (NICB)

## COMPETITOR RATE FILINGS/MANUALS

Companies may be required to submit rate filings when changing rates or rating structures. These filings typically consist of actuarial justification for the proposed rate changes and the manual pages needed to rate a policy. Most filings contain helpful information related to overall indicated loss cost levels and trends in losses and expenses.

A company must exercise caution when relying on information obtained from a competitor's rate filing, as the competitor may have significantly different insureds, goals, expense levels, and operating procedures.

## OTHER THIRD PARTY DATA

The most commonly used types of third-party data that are not specific to insurance include:

- Economic data, such as the Consumer Price Index (CPI)
- Geo-demographic data, such as the U.S. census data, weather indices, theft indices, and average annual miles driven
- Credit data

## Data Considerations

An actuary should keep in mind the following considerations when working with data.

1. **Limited data**
   In situations where preferred data is unavailable, the actuary must rely on the available data and use actuarial judgment to compensate for data limitations.

2. **Multiple currencies**
   Data in different currencies should be translated to the currency used to conduct the analysis using the appropriate exchange rates at a common point in time.

3. **Large claims**
   Large claims could distort actuarial analyses. The specific size threshold for categorizing a claim as a large claim can vary depending on factors such as the type of insurance, the geographical area, and the specific analysis being conducted. The following factors should be considered when establishing the large claim threshold:

   - Number of claims over the threshold each year
   - Size of claim relative to policy limits
   - Size of claim relative to reinsurance limits
   - Credibility of internal data regarding large claims
   - Availability of relevant external data

4. **Terminology differences**
   An insurance term may carry different meanings when being used by different insurers, TPAs, IAs, or even different departments within the same organization. It is crucial for the actuary to accurately understand the actual definitions of the data before using it.

## Verification of Data

Actuaries are often required to verify that the data they utilize is reliable and sufficient for the intended purpose. A data review usually focuses on the following areas:

- Consistency with financial statement data
- Consistency with prior data
- Reasonableness of the data
- Clarity and accuracy of data definitions