

How to run CAS Simulator 2.0

Reserveprism

CAS Simulator 2.0 is an open source R package (cascsim.zip), helping actuaries perform transaction data fitting and claim level loss reserving. Most of the methods used in simulation follows what is set out in [Triangle-Free Reserving](http://www.actuaries.org.uk/documents/triangle-free-reserving-non-traditional-framework-estimating-reserves-and-reserve) (Parodi, 2013).

It comes with an Excel Front End to assist users with an easy to use interface. Please see related documentation on methodologies, architectures, and how to use the R package directly, etc.

This document will focus on how to use the Excel API, which is named cascsim.xlsm.

Contents

[Initial Setup 2](#_Toc519211466)

[Just perform Simulation without Claim Data 4](#_Toc519211467)

[Simulation Output 12](#_Toc519211468)

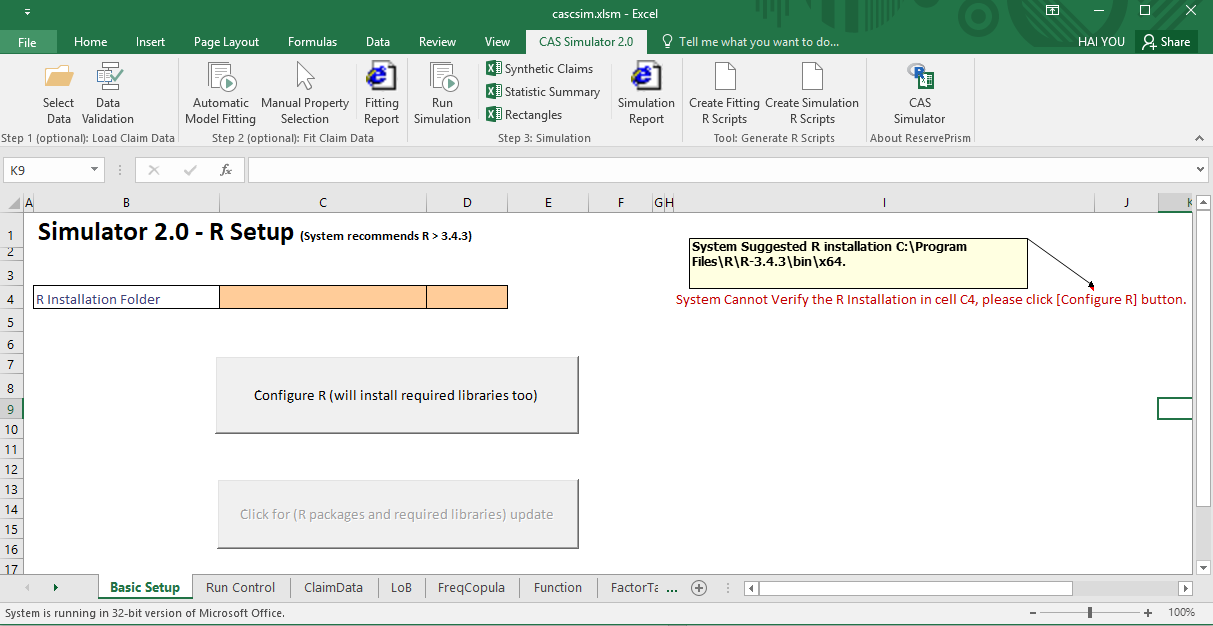
[Fit Claim Data 14](#_Toc519211469)

[Simulate with Claim Data 20](#_Toc519211470)

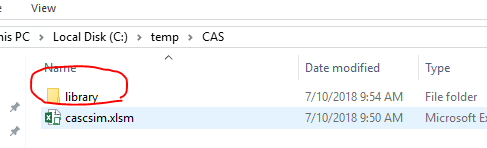
# Initial Setup

1. Download <http://www.ReservePrism.com/Simulator2/cascsim.xlsm>
2. If not yet, download and install R (recommend >=3.4.3)
3. Open the cascsim.xlsm. It will perform the following tasks for you

For first time usage, it will detect your R location, but ask you to click **[Configure R]** button and pick the suggested R location in the comment area.

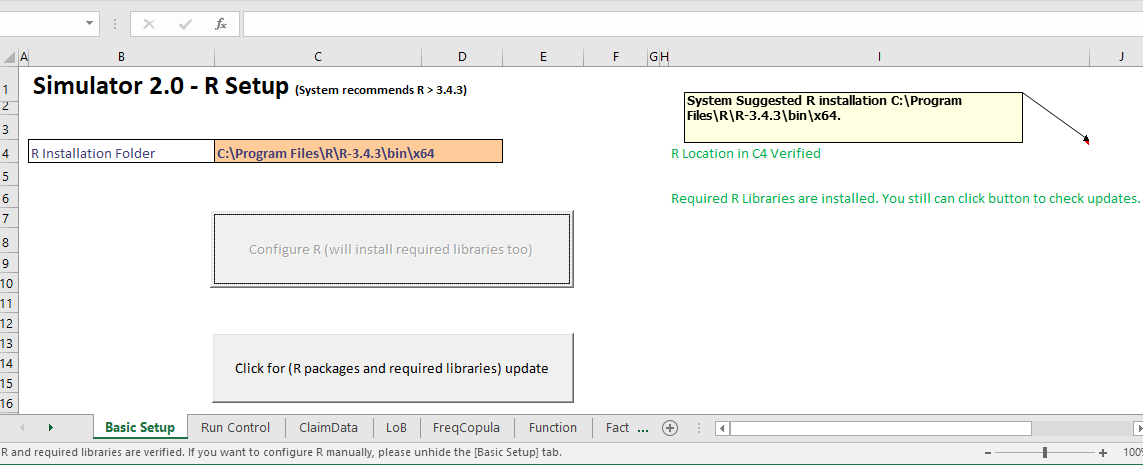


All required R packages, as well as the Simulator 2.0 package (cascsim.zip), are then installed in library folder where the Excel resides.



1. That is it! You are done with the setup

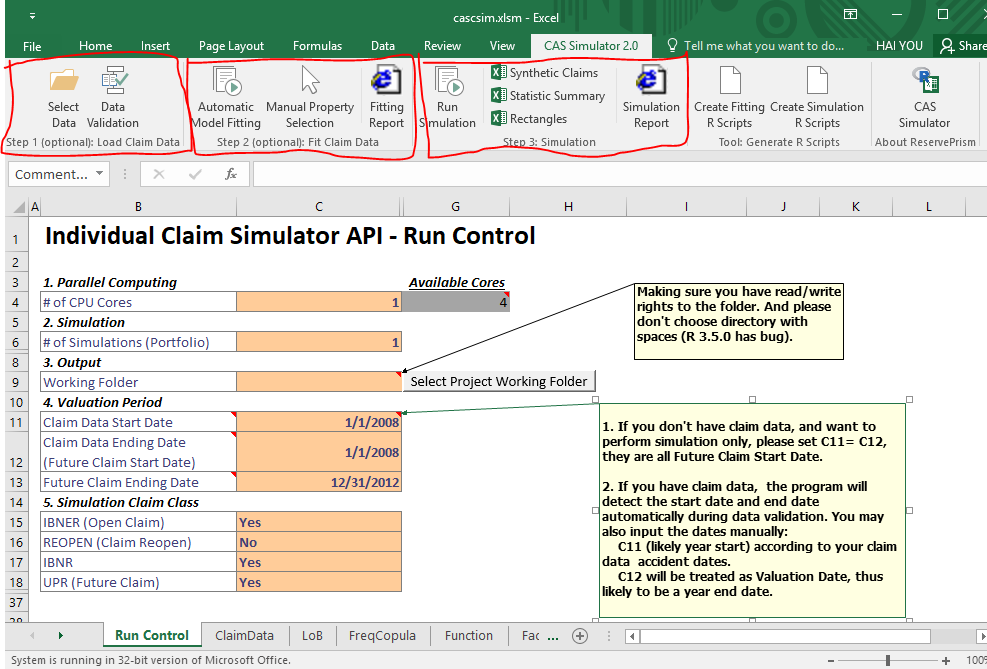
The **[Basic Setup]** tab will look like this after the successful setup.



The **[Basic Setup]** tab will be hidden in the future, yet, you can unhide it at any time and click the **[Click for update]** button to refresh all R libraries.

# Just perform Simulation without Claim Data

1. Let us take a quick look at Simulator 2.0 menu (or ribbon) items.



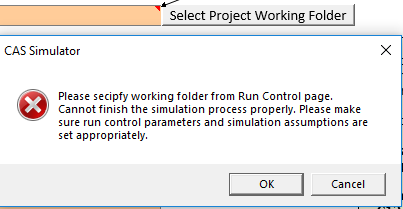
You will notice the ribbon on top of Excel for CAS Simulator 2.0.

It has Step 1, (optional) Load Claim Data and Step 2, (optional) Fit Claim Data. We will follow them in next chapter, because we believe it might be difficult for users to get transactional data. So let us start with synthetic claims simulation, which is Step 3, Simulation.

1. **[Run Control]** tab is very important.
2. Cell [C6], # of Simulations,

It tells the program how many iterations you want to simulate. For dry run, set it to 1 to save time. If you want to have a better distribution of the squared triangle for real loss reserving, set it to 100 or more, which yet will take much longer time to finish.

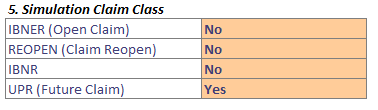
1. Cell [C9], Working Folder

Make sure you have read/write rights to the folder. Please don't choose directory with spaces (R 3.5.0 has bug). If you leave this cell blank, you will get an error notice sooner or later.

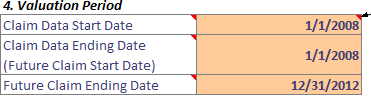
You can manage multiple working folders for different loss reserving projects, but each time you have to choose one folder only.

1. Valuation Period

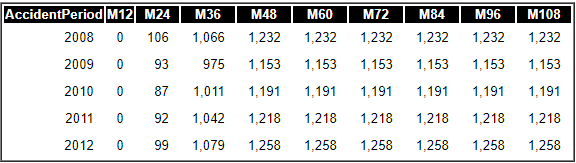
Without claim input, open claims and closed claims information are not available. The system will assume the simulation is for “future” claims. Therefore, UPR is selected by default. .



Then, just as a demo, you may set the dates as follows. In simulation without claim data, making sure to set both cell C11 and cell C12 as the starting accident date . (Note that this setup is for a simulation without historical claim data. The setup will be different when claim data is available, which is explained in a later section)



And then the rectangle (reported count) output will end up like the following (please notice the Accident Period range).



1. **[LOB]** tab contains line and type level properties.

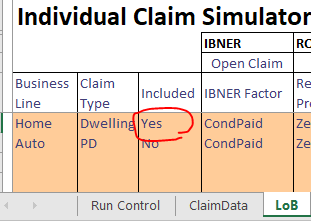
To perform Simulation only (without claim data input and fitting), you have to manually enter various distribution/index/factor values. There are default values for user to start with.

In next chapter, Claim Fitting, most properties in this tab will be automatically fitted and filled (and you can also manually choose different values)

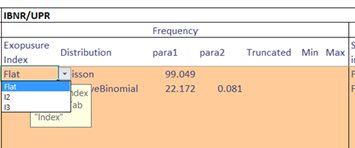
But first, let us study the **[LOB]** tab

1. Which business line(s)/claim type(s) will be included in the simulation?

When setting the [Included] column value to [Yes] for a business line/claim type, the simulation will include that business line/claim type. A value of [No] means the business line/claim type will be ignored in the simulation.



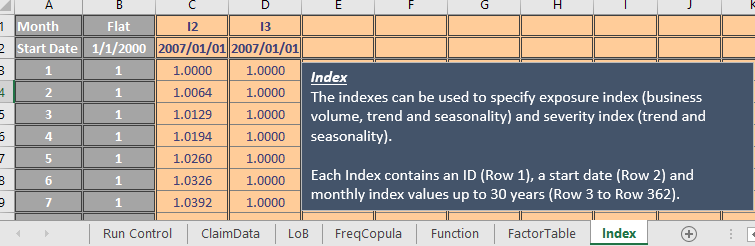
1. For pure simulation without data, we care about future claims (UPR), so let us start with frequency



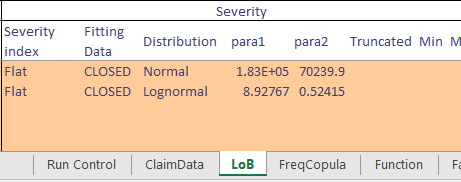
Exposure index reflects the changes in exposure by time. You can choose an index table specified in **[Index]** tab. (When simulating, detrended frequency will be simulated first and then the exposure index is applied)

Detrended frequency is set as monthly distribution. You can specify the minimum and maximum value for any distribution by set “Truncated” to “Yes” and input “Min” and/or “Max” value. The example above will give us Poisson (100) monthly, which in turn is annual frequency of Poisson (1200).

To set an index in **[Index]** tab, you can give index a name such as flat, I2, I3, etc, provide its starting month, such as “2007-01-01”, and then input monthly index value.



1. Severity Distribution

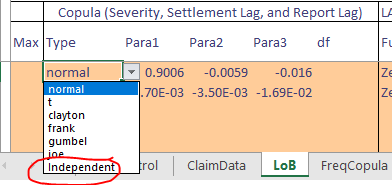


If the severity is believed to have a time-dependent trend, you can also select from the **[index]** tab that reflects the trend.

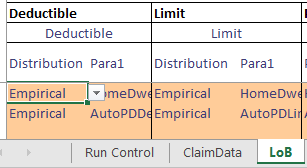
(When fitting after having input claim data, under “Fitting Data”, you can choose whether to use CLOSED claim or ALL claim (closed + open) to fit severity distribution. But since this chapter is for a simulation without historical data, it does not matter)

1. Dependency among severity, settlement lag and report lag

You can set the dependence among severity, settlement lag and report lag by a copula (normal, t, clayton, rank, gumbel, or joe copula), or you can assume they are independent. For example, for a Normal copula, Para1 is the correlation coefficient between severity and settlement lag, Para2 is that between severity and report lag, and Para 3 is that between settlement lag and report lag. df is the degree of freedom used in t copula.



1. Deductible and Limit

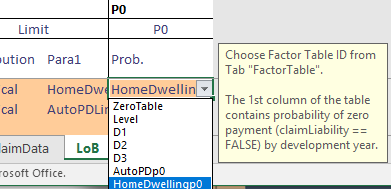


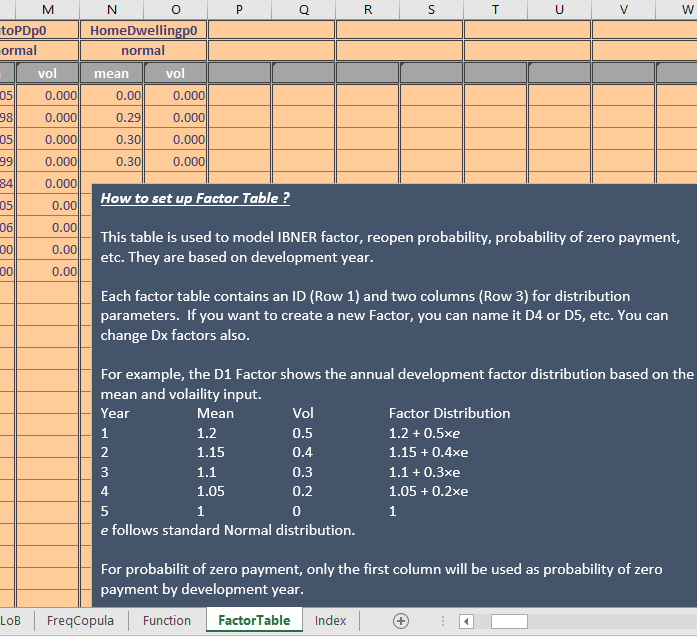
Deductible and limit can be set as

* + - “None”: no deductible or limit.
    - “Fixed”: a fixed number with value specified in Para1.
    - “Empirical”: an empirical distribution fitted from claim data with distribution name specified by Para1.

1. P0: probability of zero payment

The probability of zero payment claims is specified by development year dependent factor table set in **[FactorTable]** tab.

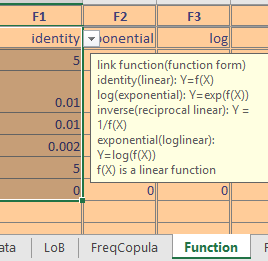




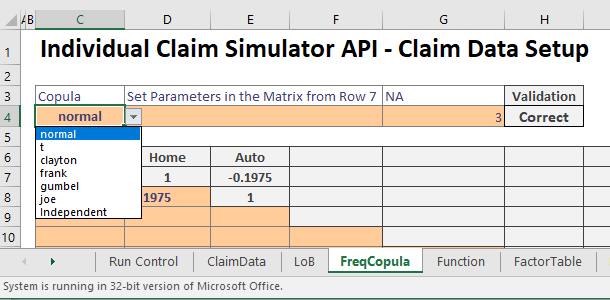
1. LAE: loss adjustment expense.

LAE can be modeled either as a function or a factor table. For example, if we think LAE = 5 + 0.01 × incurred loss + 0.01 × develop. year + 0.002 × outstanding ratio + 5 × *e* , we can create a function like “F1” in **[Function]** tab and set the LAE to “F1”. Here is a random number following standard Normal distribution.

The assumption is only applied to future claims. For open claims and closed claims where existing LAE data is available, simulated LAE is developed using the loss development factor.

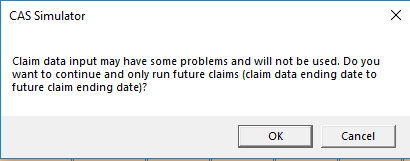


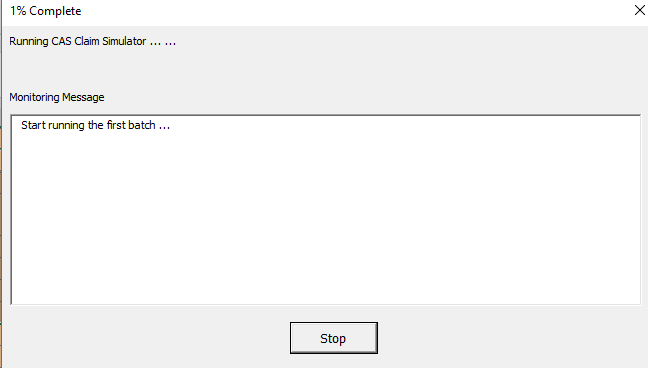
1. Tab **[FreqCopula]** contains frequency copula among business lines.



# Simulation Output

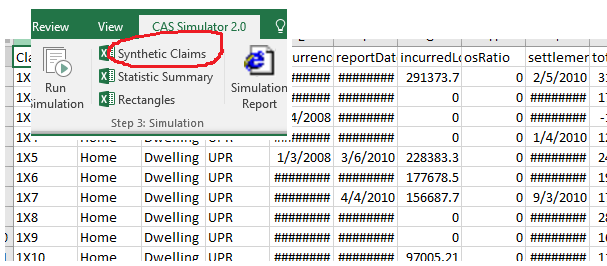
With the assumptions set from previous chapter, let us start simulation by clicking the **[Run Simulation]** Ribbon Button at top. Ignore the following warning since we purposely want simulation without historical claim only,





Once succeeded, let us see what we have

1. Synthetic claims. The Simulator 2.0 only handles single payment process, and there is no case reserving module built in yet, there is no transaction data available. Summary and rectangles based on synthetic claims are also generated and accessible.

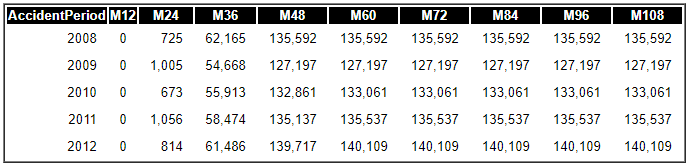


1. Simulation Report

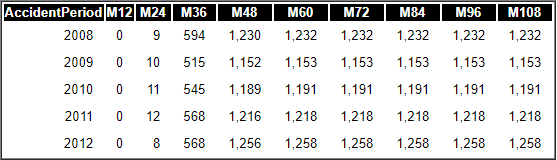
A web report is generated, containing all percentiles and statistical analysis

For example, following rectangle is Average Simulated Incurred Loss Rectangle (Thousands).

Please note, this document shows a very long report lag for this demo.

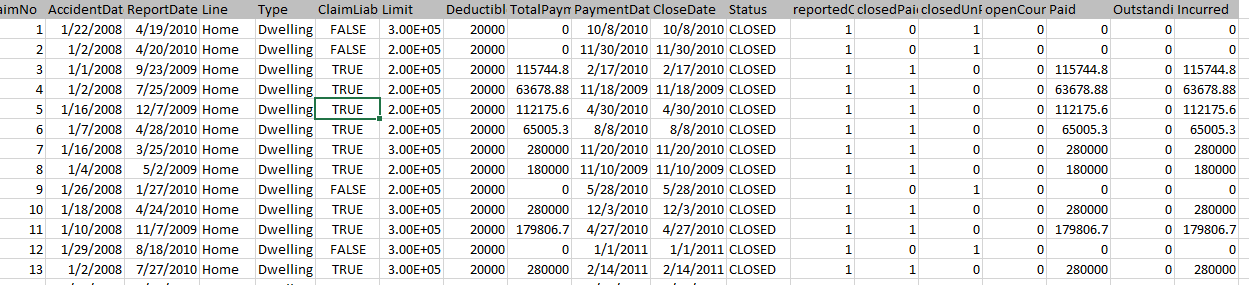


The following is Average Simulated Closed Count Rectangle



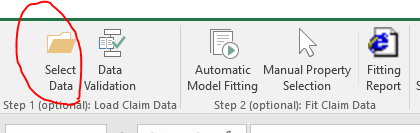
# Fit Claim Data

Simulator 2.0 is currently built in with only a Single Payment Transaction Model. By default, it only accepts loss information at individual claim level. If your claim data have transaction details, at this stage, please aggregate all the transactions (payments) for each claim(just like your year-end claim format)

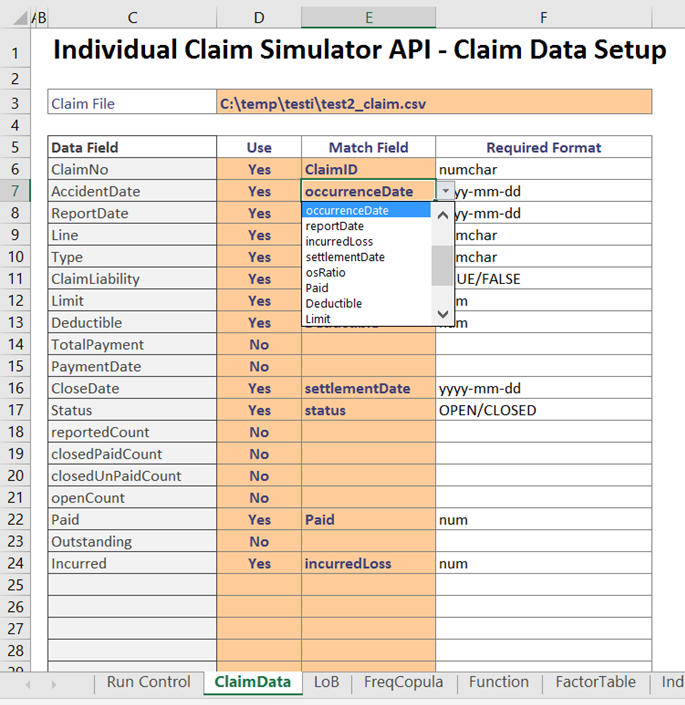


Once claim file (csv format) is provided, system will fit it to the **[LoB]** tab properties as described in previous chapter, thus simulation becomes a realistic loss reserving analysis. (Simulation will help loss reserving analysis by providing close-to-real synthetic claim transactions and rectangles)

1. Select the claim data file by pressing ribbon button **[Select Data]**.



1. All the data fields in the csv file will be populated (column C) in **[ClaimData]** tab. User needs to determine whether or not each data field should be used (column D) and its matching data field name (column E), for Simulator to fit from.

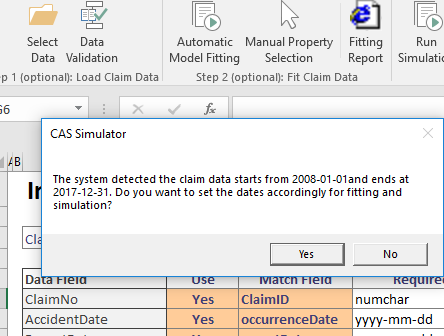


The following variables are mandatory:

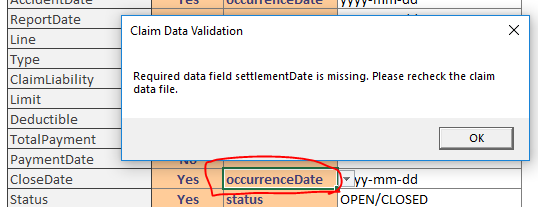
* ClaimID
* LoB
* Type
* status
* occurrenceDate
* reportDate
* incurredLoss
* settlementDate

Other standard variables can be provided as well, such as: osRatio, Paid, Deductible, Limit, totalLoss, LEA, claimLiability, etc. If you have non-standard variables that will be used for GLM functions, you can include them in the analysis as well.

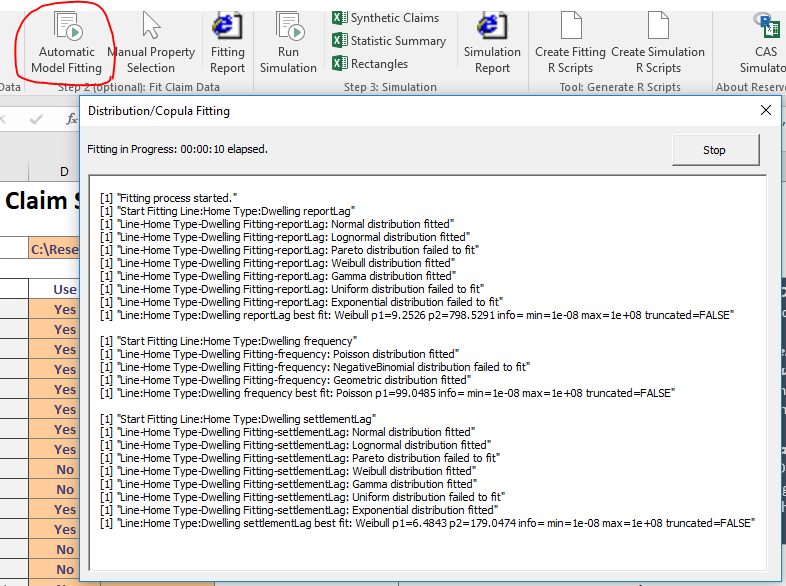
1. By pressing ribbon button **[Data Validation]**, the program will perform various pre-fitting data validations.



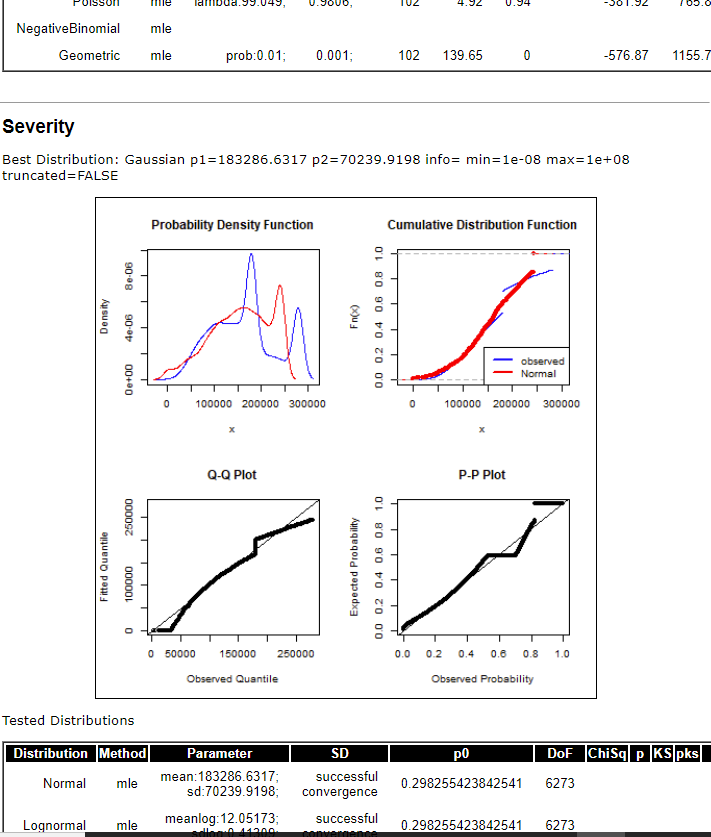
For example, if you mistakenly match a wrong date, you will be notified.



1. Once ready, press the ribbon button **[Automatic Model Fitting]**, the program will start fitting with monitoring messages. To fit any property, the system will pick the best statistical fit from various distribution assumptions.



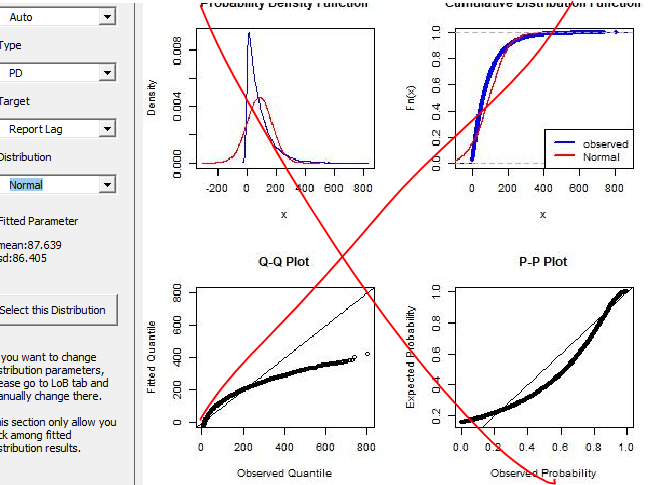
Once the fitting is done, most appropriate assumptions (distribution, copula, and P0) are populated in tab **[LoB]**. A fitting report can also be viewed by pressing ribbon button **[Fitting Report]**.

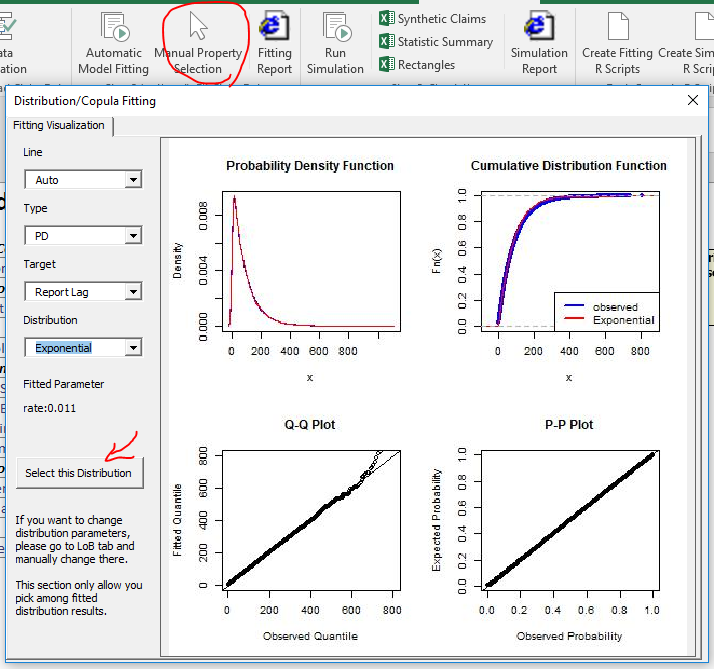


1. You can also review each individual fitting by pressing ribbon button **[Manual Property Selection]**. If you find a fitting is appropriate, you can press button **[Select this Distribution]** to replace the assumption in tab **[LoB]**.

The following picture shows an example that, after automatic fitting, you can manually choose Report Lag Distribution property. (following Report Lag property can be chosen as):

Simply from Q-Q plot observation, you find out that Normal Distribution is not be a good pick. Instead, an Exponential Distribution should be a good fit.

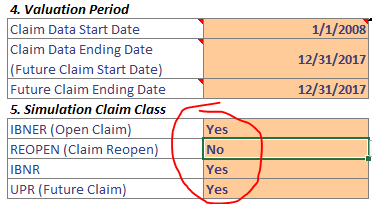




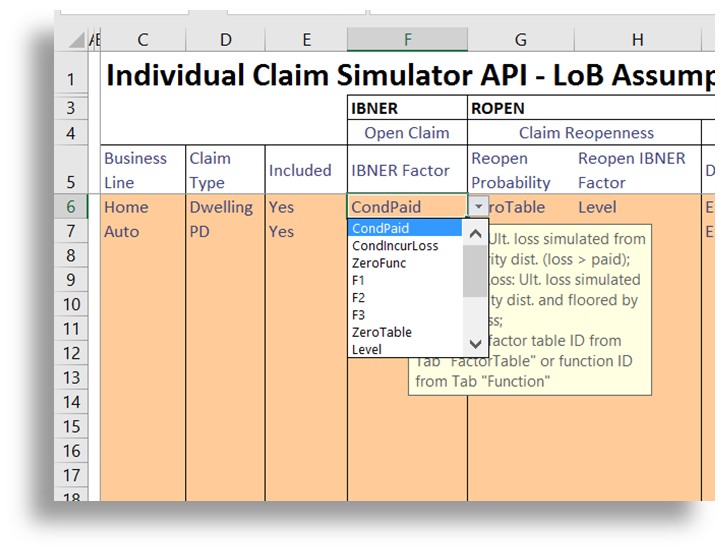
# SimulatION with Claim Data

Once claim data fitting is finished, let us do real time loss reserving by simulating open-claims development (IBNER), IBNR claims and closed claim reopening. We need to set a few more assumptions in **[Run Control]** tab.

1. Specify the classes of claims to simulate in tab **[Run Control]**.



1. Set open-claim loss development assumption (IBNER Factor) in tab **[LoB]** to one of two options:

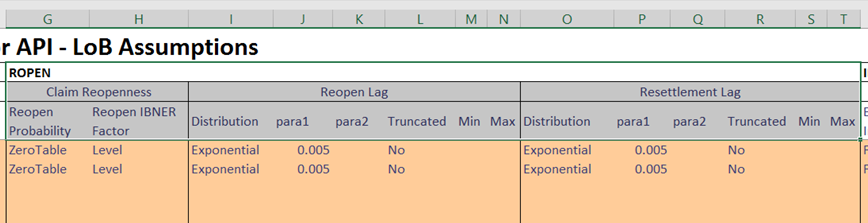


CondPaid: Ultimate loss simulated from the selected severity distribution, conditional on the loss being greater than the claim’s paid amount;

CondIncurLoss: Ultimate loss simulated from the selected severity distribution and floored by the claim’s incurred loss amount;

Alternatively, you can choose to model the loss development factors by selecting a factor table from tab **[FactorTable]** or a function from tab **[Function]**.

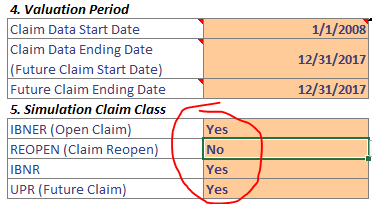
1. If you want to model closed claim reopening, you can set reopen probability as a function or a factor table. Reopen claim loss development factors can be set the same way as you set IBNER factors for open claims. You then set reopen lag and resettlement lag as distributions.



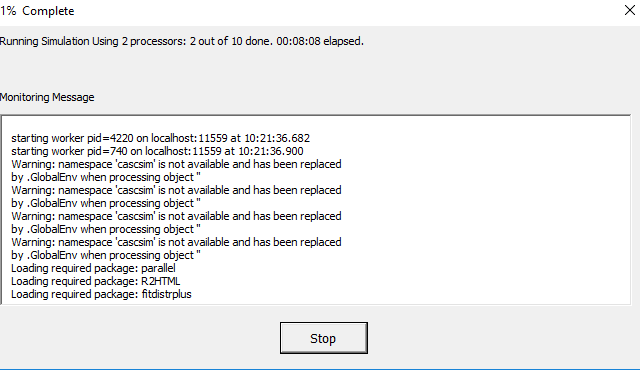
Other assumption setting and simulation process are the same as in the case of simulation without claim data.

1. Let us take a quick look on final simulated rectangles and percentiles.

Again, this is the setting from the beginning of this chapter.



This is from a 10-iterations simulation, for demonstration purposes only. At this stage, only thing we need is patience, to let it finish. (The Excel API and R is not doing real-time synchronization, there is a time gap for Excel to read R logs)



For example, this is part of the final statistical analysis, along with all other summary reports.

