Pattern break detection

Othman El Hammouchi

Introduction

Stochastic claims reserving methods in non-life insurance are based on models whose underlying assumptions are difficult verify using classical statistical inference given the limited amount of data available. A possible way to remedy this, which we will explore in this notebook, is through the use of bootstrap simulation. We focus on Mack's method, which is one the most popular reserving techniques. It makes the following assumptions:

$$\mathbb{E}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathbb{E}[C_{ij} \parallel C_{i,j-1}] = f_{j-1} C_{i,j-1} \tag{1}$$
$$\mathrm{Var}[C_{ij} \parallel C_{i,j-1}, \ldots, C_{i1}] = \mathrm{Var}[C_{ij} \parallel C_{i,j-1}] = \sigma^2_{j-1} C_{i,j-1} \tag{2}$$

, for $i \in \{0, \ldots, I\}$ and

$$\{C_{i,0}, \ldots, C_{i,I-i}\}, \{C_{i',0}, \ldots, C_{i',I-i'}\} \text{ independent for } i \neq i' \,. \tag{3}$$

We start by gauging the sensitivity of the simulated reserve to deviations from these assumptions. In other words, starting from a sythetic triangle which is constructed to satisfy the assumptions perfectly, we apply various violations and study the effect on the predictive distribution of the reserve generated by a bootstrapping procedure. If such violations turn out to have a significant effect, this suggests a method for detecting problematic observations in real claims data.

Perturbing a single observation

The first violation we consider is of the mean assumption. We can repeat the simulation for every possible point and a range of perturbations between 0.5 and 1.5.

triangle conforming to Mack's assumptions. For added realism, we borrow the development factors and σ_j-parameters from a real dataset, the UKMotor data from the `ChainLadder` package.

```
## Warning in Mack.S.E(CL[["Models"]], FullTriangle, est.sigma = est.sigma, : 'loglinear' model to estim
## p-value > 5.
##  est.sigma will be overwritten to 'Mack'.
##  Mack's estimation method will be used instead.
```

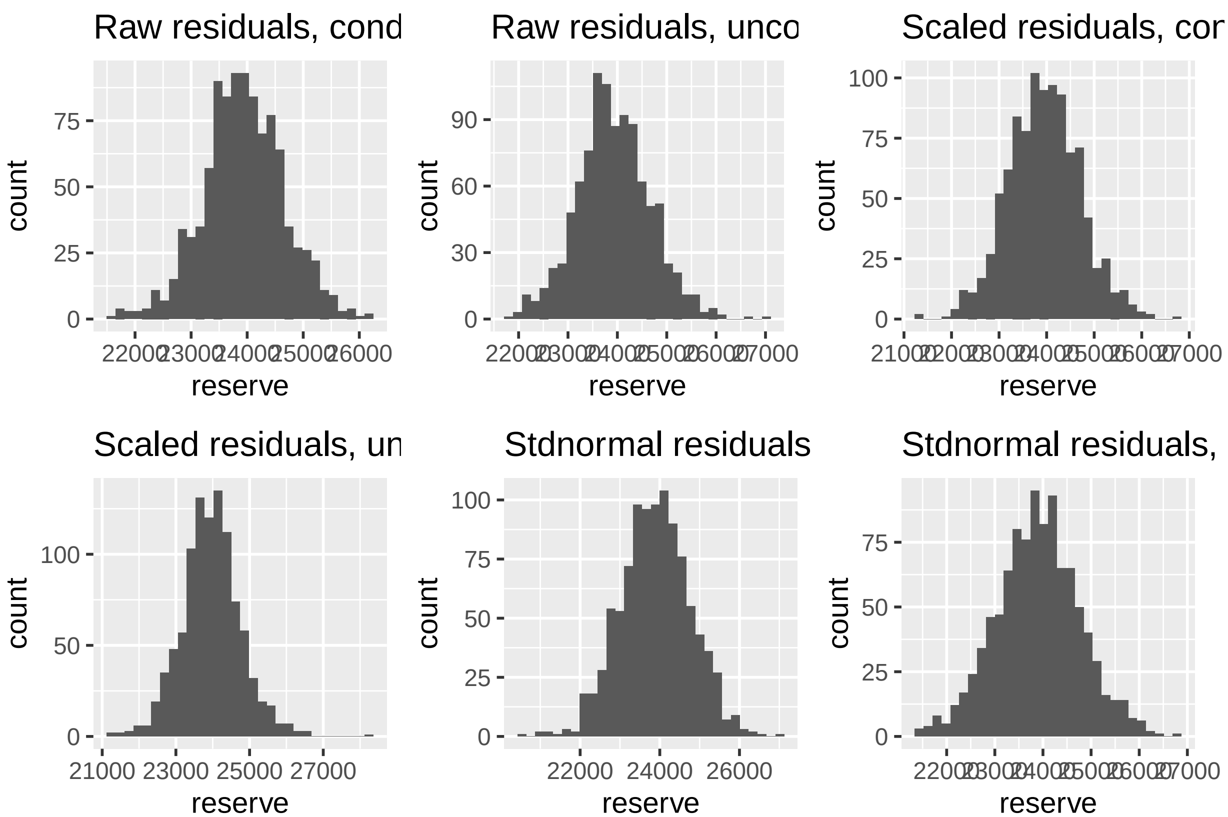
3764.504	7001.556	8934.617	10316.745	11474.44	12053.85	12389.4
5785.374	10868.329	13563.978	15469.333	17060.13	17957.60	NA
3387.945	6423.402	8200.029	9649.186	10668.86	NA	NA
4591.844	8910.815	11238.649	13189.326	NA	NA	NA
6082.420	11313.911	13732.489	NA	NA	NA	NA
3980.601	7327.118	NA	NA	NA	NA	NA
4099.336	NA	NA	NA	NA	NA	NA

As expected, the Mack estimator performs very well in this case:

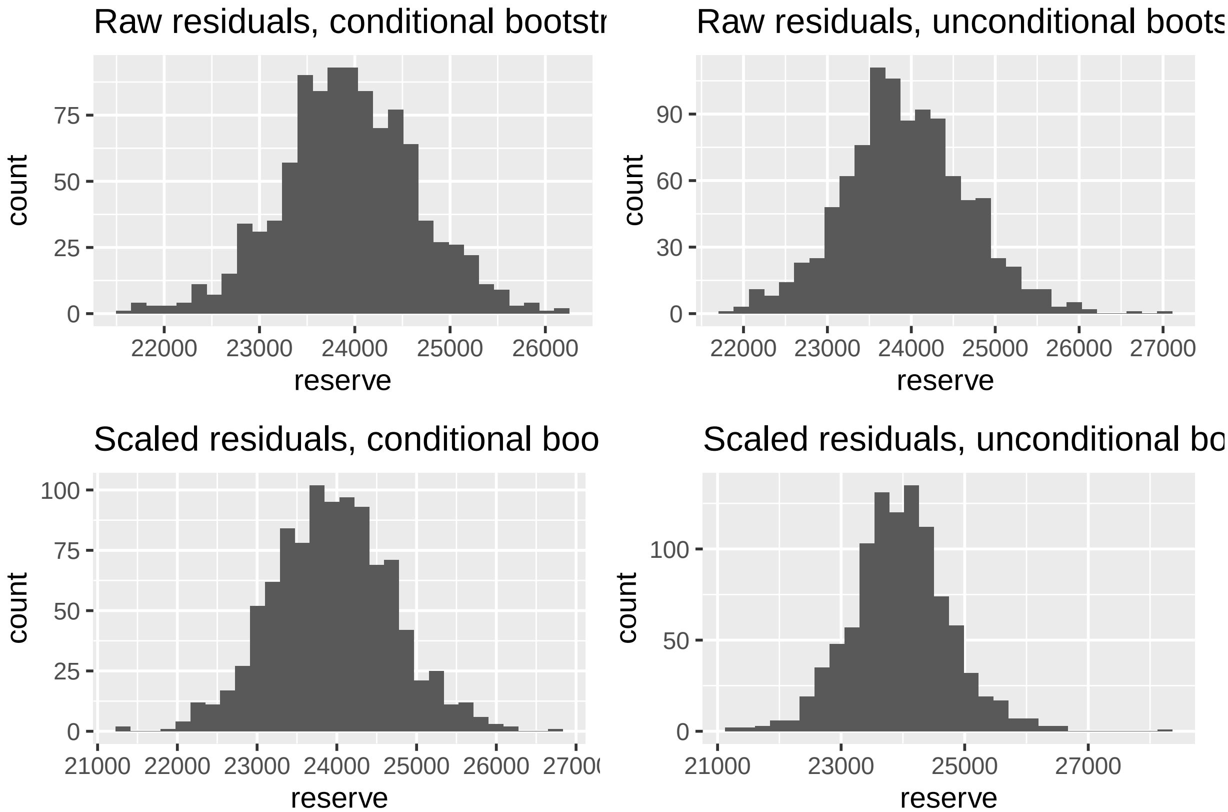
```
## [1] 1.878945 1.250500 1.159460 1.106339 1.051758 1.027838 1.000000
```

We will study different kinds of bootstrapping procedures. Let's start by comparing them.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



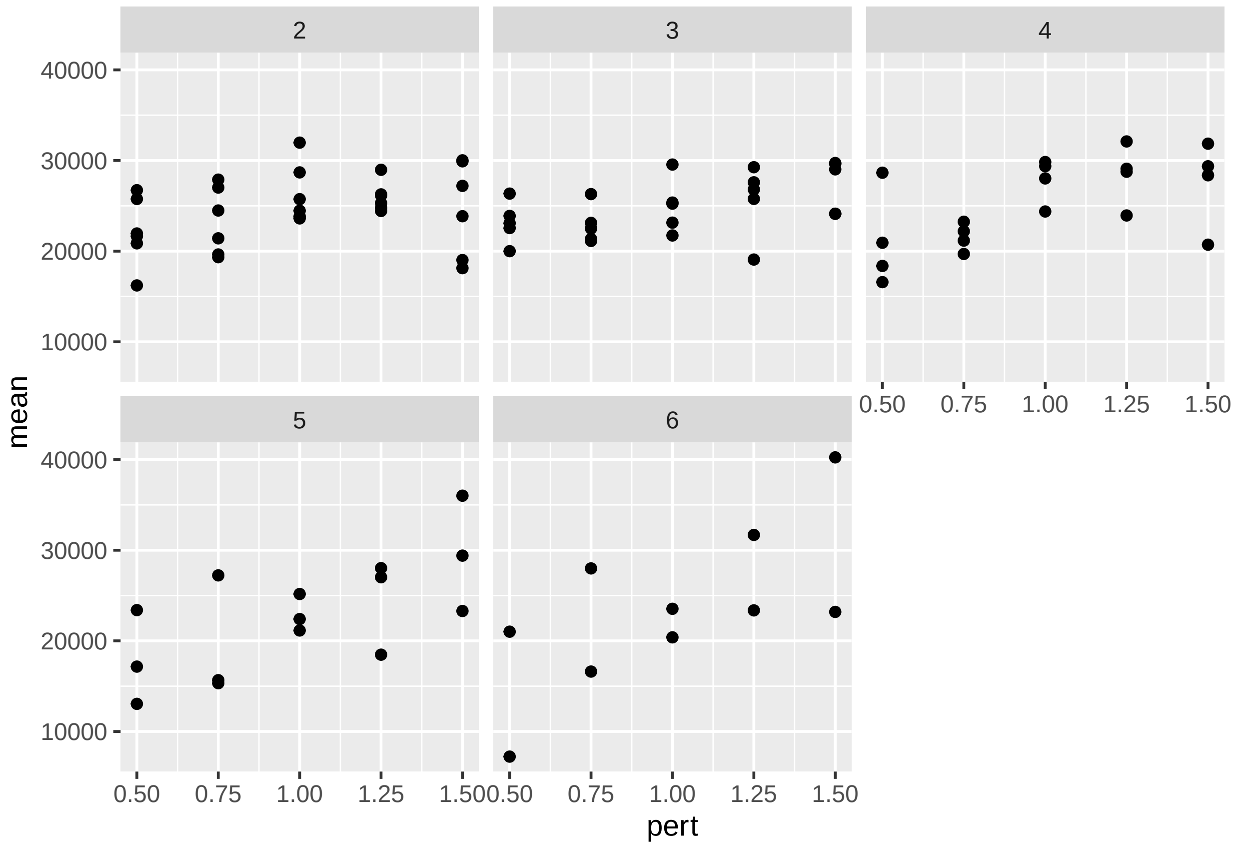

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

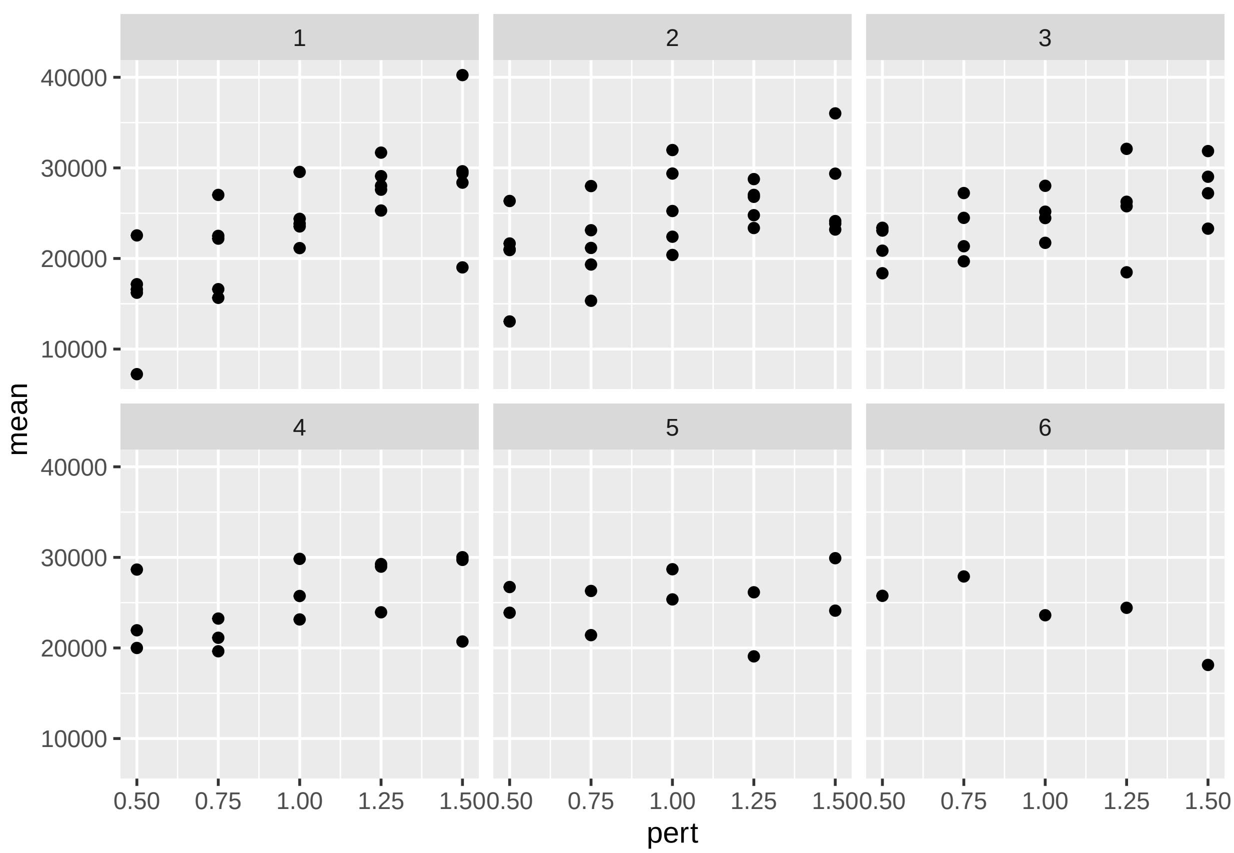


Sequential code:

Parallel code:

Now let's take a look at the bootstrap mean as a function of the perturbation for different rows and columns.


2
3
4
5
6
mean
pert
40000
30000
20000
10000
0.50
0.75
1.00
1.25
1.50


1
2
3
4
5
6
mean
pert
40000
30000
20000
10000
0.50
0.75
1.00
1.25
1.50

Now we want to go in the reverse direction: if we start from a perturbed triangle, does removal of the outlying observation create a significant difference in the reserve?

```
## # A tibble: 1,000 x 5
##    indivDevFacBoot devFacBoot sigmaBoot triangleBoot  reserve
##    <list>          <list>     <list>    <list>          <dbl>
##  1 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  22066.
##  2 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  26800.
##  3 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  23744.
##  4 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  23256.
##  5 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  24204.
##  6 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  30791.
##  7 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  20158.
##  8 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  25493.
##  9 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  33822.
## 10 <list [6]>      <dbl [6]>  <dbl [6]> <dbl [7 x 7]>  25626.
## # ... with 990 more rows
```

Gamma distribution

The problem with using the above normal model is that it allows for negative samples during the bootstrap simulation. To avoid this, we can use a different distribution which still respects Mack's assumptions. If we take for instance $C_{ij} \sim \Gamma(\alpha, \beta)$, then we must choose α, β to satisfy

$$\begin{cases} \frac{\alpha}{\beta} = f_{j-1} C_{i,j-1} \\ \frac{\alpha}{\beta^2} = \sigma_{j-1}^2 C_{i,j-1} \,, \end{cases}$$

giving the values

$$\alpha = \frac{f_{j-1}^2 C_{i,j-1}}{\sigma_{j-1}^2} \beta = \frac{f_{j-1}}{\sigma_{j-1}^2} \,,$$

for the distribution parameters.