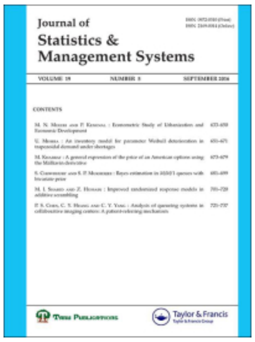

Taylor & Francis
Taylor & Francis Group

Journal of Statistics and Management Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsms20

Backtesting the Bayesian Bornhuetter-Ferguson method against traditional approaches in claims reserving

Matteo Crisafulli & Gian Paolo Clemente

To cite this article: Matteo Crisafulli & Gian Paolo Clemente (2022) Backtesting the Bayesian Bornhuetter-Ferguson method against traditional approaches in claims reserving, Journal of Statistics and Management Systems, 25:8, 1919-1943, DOI: 10.1080/09720510.2021.1995216

To link to this article: https://doi.org/10.1080/09720510.2021.1995216

Published online: 26 Jul 2022.

Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 25 (2022), No. 8-A, pp. 1919–1943
DOI : 10.1080/09720510.2021.1995216

Backtesting the Bayesian Bornhuetter-Ferguson method against traditional approaches in claims reserving

Matteo Crisafulli *
Department of Statistical Sciences
Università degli Studi di Roma-La Sapienza
Roma
Italy

Gian Paolo Clemente †
Department of Mathematics for Economic,
Financial and Actuarial Sciences
Università Cattolica del Sacro Cuore
Largo Gemelli 1
Milano 20123
Italy

Abstract

Evaluation of claims reserve is a paramount process for non-life insurance company. To this end, several deterministic and stochastic methodologies have been provided in the literature. Therefore, the validation of the models on actual data and the comparison of these models appropriateness is nowadays a crucial question. We focus here on different Bornhuetter-Ferguson methodologies and we backtest the behavior of these models using the well-known dataset made available in [22]. The aim is to test both the ability of different models to well predict future losses as well as to evaluate the effects of different priors on the results. Additionally, we test the uncertainty of the predictions by comparing the coefficient of variation.

Subject Classification: *62P05.*

Keywords: *Claims reserves, Bornhuetter-Ferguson methodologies, Backtesting, Markov-Chain-Monte Carlo.*

* *E-mail:* `matteo.crisafulli@uniroma1.it (Corresponding Author)`
† *E-mail:* `gianpaolo.clemente@unicatt.it`

1. Introduction

Best estimate of unpaid claims is probably the most important liability for non-life insurance companies. Earnings reports, financial statements and management decisions are affected by this evaluation. Traditionally, the evaluation of claims reserve focused on deterministic methodology (see [9]). Over the last twenty years, stochastic methodologies allowed to include percentiles and confidence regions at accident, development or calendar years levels. In particular, rather than trying to provide a specific estimate, the unpaid claim analysis process has focused on understanding the variability around the estimate by identifying a range of reasonable estimates using different methods and assumptions.

The validation of the models on actual data and the comparison of these models' appropriateness is nowadays a crucial question. Comparison and selection of optimal models have been recently explored in the literature. An investigation about bootstrap models has been developed in [20]. A case study is performed in [35] in order to analyze the accounting year effects in the triangles. [29] and [28] provide other comparisons based on QQ-plots and PP-plots. A specific focus has been devoted to the validation of methods in [16]. Three methods, the double chain ladder, the Bornhuetter-Ferguson and the incurred double chain ladder methods are compared through two real data sets from property and casualty insurers. Supported by real-life claims data, the authors in [30] compare three models with different residual adjustments using the Dawid–Sebastiani scoring rule. Alternative stochastic claims reserving methods are compared in [17] by means of a wide range of goodness-of-fit measures.

In this paper, we focus on different Bornhuetter-Ferguson models provided in the literature and we backtest their behavior on the database published in [22]. The aim is to test both the ability of different models to well predict future losses as well as to evaluate the effects of different priors on the results. The performances of these models are then compared with the classical bootstrap over-dispersed Poisson model.

Additionally, we test the uncertainty of the predictions by comparing the coefficient of variation of different models. The paper is structured as follows. In Section 2, we describe the methodological aspects of the Bayesian Bornhuetter-Ferguson model. In particular, we focus on the alternative priors that can be introduced and will be tested in the numerical section. In Section 3 we describe the dataset and we provide a detailed description of the data selection process applied to assure consistency in the model comparisons. Additionally, a synthetic description of the compared models is provided. In Section 4, we summarize the alternative metrics used to analyze models' performance and determine an order of appropriateness. Additionally, main results are reported with a related discussion. Conclusions follow.

2. Bayesian Bornhuetter-Ferguson model

We mainly describe in this section the Bayesian Bornhuetter-Ferguson (BBF) model provided in [8] that will be tested in Section 4.

We consider a run-off triangle with rows $i = 0,\dots,n$ and columns $j = 0,\dots,n$ representing accident and development years, respectively. This structure is peculiar in the actuarial context and is used in order to project elements to their ultimate value, according to a past observed pattern, opportunely weighted. This run-off triangle can be divided in two parts, the upper one, representing observed elements, and the lower one, representing what has to be estimated by the model. In practice, claims reserving approach are usually based on this structure, and try to estimate future realizations by means of the set of past observations. We denote the upper triangle of incremental payments $P_{i,j}$ using the set $D_n = \{P_{i,j} : i + j \le n\}$.

We start by reporting main assumptions of the BBF model:

(1) There exists $\mu_0,\dots,\mu_n,\gamma_0,\gamma_n,\phi$ independent random variables with joint density $u(\cdot)$

(2) Parameters μ_i are Γ-distributed, with mean $m_i > 0$ and shape parameter $a_i > 0$:

$$\mu_i \sim \Gamma\left(a_i, \frac{a_i}{m_i}\right)$$

(3) Parameters γ_j are Γ-distributed, with mean $c_j > 0$ and shape parameter $b > 0$:

$$\gamma_j \sim \Gamma\left(b, \frac{b}{c_j}\right)$$

(4) Parameter $\phi > 0$ is a constant

(5) Conditionally given $\Theta = (\mu_0,\dots,\mu_n,\gamma_0,\gamma_n,\phi)$, the incremental payments $P_{i,j}$ are independent random variables with:

$$\left.\frac{P_{i,j}}{\phi}\right|_{\Theta} \sim Poi\left(\frac{\mu_i \cdot \gamma_j}{\phi}\right)$$

The posterior density of Bayesian over-dispersed Poisson (ODP) model can then be expressed, given D_n, as:

$$u(\theta \mid D_n) = \frac{f(\boldsymbol{p},\theta)}{f(\boldsymbol{p})} = \prod_{i+j\le n}\left(\exp\left\{-\frac{\mu_i \cdot \gamma_j}{\phi}\right\}\frac{\left(\frac{\mu_i \cdot \gamma_j}{\phi}\right)^{\frac{P_{i,j}}{\phi}}}{\Gamma\left(1+\frac{P_{i,j}}{\phi}\right)}\frac{1}{\phi}\right)$$

$$\times\prod_{i=0}^{n}\frac{\left(\frac{a_i}{m_i}\right)^{a_i}}{\Gamma(a_i)}\mu_i^{a_i-1}\exp\left\{-\frac{a_i}{m_i}\mu_i\right\}$$

$$\times\prod_{j=0}^{n}\frac{\left(\frac{b}{c_j}\right)^{b}}{\Gamma(b)}\gamma_j^{b-1}\exp\left\{-\frac{b}{c_j}\gamma_j\right\}u(\phi)\cdot\frac{1}{f(\boldsymbol{p})}$$

Since the model is embedded in a Bayesian framework, one of the most important point consists in the choice of the amount of information contained in the priors. Given the assumption of gamma distributions for μ_i and γ_j, the informativity can be expressed by the shape parameters a_i and *b*, or analogously by the coefficient of variations (CV) $1/a_i$ and $1/b$. Using the coefficient of variation of the prior distribution, it is possible to divide the priors in [32]:

$$CV_u(\cdot)\begin{cases} =0 & \text{strong prior} \\ \in(0,1) & \text{informative prior} \\ \in[1,\infty) & \text{vague prior} \\ =\infty & \text{non-informative prior} \end{cases}$$

It is worth pointing out two extreme situations: the strong prior case, where the distribution is degenerate on its mean and the non-informative prior case, where the distribution is spread with same probability on its whole domain.

Considering the claims development parameter γ_j, it is possible to assume a non-informative prior and obtain a closed form (see [8] for the proof) of the maximum a-posteriori (MAP) predictor for the strong and non-informative μ_i priors. As regard to the strong prior case, it is possible to obtain in a closed form the estimate of claims reserve according to the minimum mean squared error (MMSE) predictor, which corresponds to the reserve provided in [15] with raw claims development parameters.

Except for the cases mentioned above, the structure of the model does not permit to arrive at a closed solution, but requires the application of a Markov-Chain Monte Carlo (MCMC) approach (see, e.g., [19] [12]) for estimating the posterior distribution of the parameters and the (predictive posterior distribution of) claims reserve. Since each row and column parameter is distributed (conditionally) as an independent Gamma distribution with specific parameters, it is possible to apply he Gibbs sampler procedure to estimate the posterior distribution of each parameter.

Hence from the empirical sample $(\Theta^{(s)})_{s>S}$ obtained by the Gibbs sampler, we can simulate the incremental payments of the lower triangle:

$$\frac{P_{i,j}^{(s)}}{\phi} \sim Poi\left(\frac{\mu_i^{(s)} \cdot \gamma_j^{(s)}}{\phi}\right) \tag{1}$$

From this result, it is then possible to estimate all the elements we are interested in the run-off triangle. In particular, we can easily derive the sample of claims reserve for all the accident year $(R_i^{(s)})_{s>S}$, where each element is computed by means of:

$$R_i^{(s)} = \sum_{j=n-i+1}^{n} P_{i,j}^{(s)} \tag{2}$$

This sample provides the empirical posterior distribution of R_i, from which we can obtain the mean and MSEP of claims reserve.

3. Preliminaries

3.1 *Objective of the analysis*

The empirical analysis has the objective of studying the application of the model described in previous section on a real dataset of insurance companies and to compare the results with a classical Bootstrap Over-dispersed Poisson (ODP) methodology (see [7] and [6]).

In the first part of the analysis we study the predictive capacity of the model, by comparing a series of predictive statistics computed as function of the model outcomes and considering the observed payments.

Then we study the variability of the estimation, by analyzing the pattern of coefficient of variation of estimated claims reserve for companies with different sizes.

3.2 *Data description*

The analysis described in the following section is based on the information coming from a dataset containing different companies' relevant information. This dataset is provided in [20] and is based on *Schedule P* of American companies published by *National Association of Insurance Commissioners* (NAIC), the U.S. standard-setting and regulatory support organization created and governed by the chief insurance regulators from the 50 states, the District of Columbia and five U.S. territories.

More specifically we have six separate datasets, one for each line of business (LoB), with data for the companies operating in that business. These lines correspond to homogeneous segments of insurance portfolios based on coverage types, whose definitions can be found in [24], and are typically characterized by different run-off behaviors, payment patterns and claims reserve volatility. In particular, they are *Commercial auto and truck liability and medical*, *Medical malpractice*, *Private passenger auto liability and medical*, *Product liability*, *Workers' compensation* and *Other liability*.

Table 1 summarizes the information reported in the dataset of a given LoB.

Table 1

Dataset description

	Variable name	Description
1.	GRCODE NAIC	Company code, including both single companies and groups
2.	GRNAME NAIC	Company name, including both single companies and groups
3.	AccidentYear	Accident year, ranging from 1988 to 1987
4.	DevelopmentYear	Development year, ranging from 19988 to 1987
5.	DevelopmentLag	Development lag, ranging from 0 to 9
6.	IncurrLoss	Incurred losses and allocated expenses reported at year end
7.	CumPaidLoss	Cumulative paid losses and allocated expenses at year end

Contd...

8.	BulkLoss	Bulk and IBNR reserves on net losses
9.	PostedReserve97	Posted reserves
10.	EarnedPremDIR	Premiums earned at incurral year - direct and assumed
11.	EarnedPremCeded	Premiums earned at incurral year - ceded
12.	EarnedPremNet	Premiums earned at incurral year - net
13.	Single	Entity indicator, 1 for single entity and 0 for group insurer

We have at disposal the fundamental information needed in order to apply claims reserving methods. In particular we can determine approximately the size of the company and its development over time by means of EarnedPrem variables. From the variables CumPaidLoss and IncurrLoss we can build the upper triangles used for applying the specific reserving methods and backtest the results by means of the same data, built using the information on the whole development.

3.3 *Data selection, pre-processing and simulation setting*

The dataset presented above also includes several companies showing an uncommon behavior, because of closure, merging, change of business or other events. It shall be noted that we do not have at disposal the specific information about the actual situation for each company, but we had to infer from the Schedule P data only. Therefore, in order to present an empirical analysis as much consistent as possible, we chose to select only a subset of the whole dataset, based on quality and consistency of the company data available.

We decided to follow the logic of *Data selection process* proposed in [20] and excluded from the analysis the companies showing an inconsistent dynamic over time. For instance, we removed companies with earned premiums equal to 0 or negative.

The number of companies before and after the data selection process are reported in Table 2; while the main statistics of the selected companies are reported in Table 3. Given the high volatility of the data for Medical malpractice and Products Liability, we preferred to focus only on the other four LoBs. We notice in Table 3 how the Private Passenger LoB is characterized by important differences in terms of companies' size, while the ultimate loss ratio distribution shows the lowest relative volatility.

Table 2

Lines of business and number of companies

Line of Business	N available companies	N selected companies
Commercial auto/truck liability/medical	158	50
Private passenger auto liability/medical	146	50
Workers' compensation	132	45
Other liability	239	45

Table 3

Main statistics on selected companies. For each LoB, we report the average premium and the average ultimate Loss Ratio observed. Between brackets, we display the coefficient of variation of premium and loss ratio distributions, respectively

Line of Business	Average premium (CV)	Average ultimate LR (CV)
Commercial auto/truck liability/ medical	185,117.3 (289.46%)	66.84% (24.71%)
Private passenger auto liability/ medical	3,029,476 (552.29%)	75.10% (13.62%)
Workers' compensation	293,505 (188.16%)	62.56% (17.18%)
Other liability	115,430.7 (328.73%)	52.14% (41.39%)

In order to perform the empirical analysis of the claims reserving models on the selected dataset we need to define a precise simulation setting. We choose to have $S = 100{,}000$ samples from simulations for each model. We have specified $100{,}000$ samples instead of simulations, because for the BBF models, being based on a MCMC approach, we have applied a burn-in of the first $b = 10{,}000$ simulations, in order to minimize the effect of the starting parameters. This b parameter has been chosen big enough to ensure this scope, without a specific calibration. In addition, we did not apply any thinning of the chain since, from a preliminary analysis on some sample companies, it seems that the autocorrelation was not relevant, and in addition it would have required a much larger number of simulations in order to achieve the same sample size.

An important point regarding the application of these models on the dataset is about the cases where run-off triangles showed incremental payments lower than 0 in the upper part. This circumstance can be problematic for both Bootstrap and BBF models. In particular, this is related with the framework described in Section 2 for BBF models, and similar for Bootstrap, which assumes that incremental payments are ODP distributed. This practical issue has been solved as described in Section 4 of [27].

3.4 *Models description*

In the empirical analysis we compare ODP Bootstrap with 5 BBF models, each one based on a different calibration of its priors. In particular, the characteristics of these models are reported below. We briefly list the characteristics of the model applied to the dataset described in Section 3.3

1. ODP Bootstrap

The application of the ODP Bootstrap is based on the theoretical framework described in [7] and [6].

2. BBF1

The first version of the BBF model is based on the theoretical framework described in Section 2, assuming that the prior information for the ultimate losses is based on the estimate obtained from the application of Mack BF model [14].

It means that the mean of the prior parameter μ_i is obtained as:

$$\mathbb{E}(\mu_i) = B_i \cdot \widehat{LR}_i = B_i \cdot r_i^* \cdot \hat{m}^* \quad (3)$$

where the term B_i represents the amount of earned premiums and $\widehat{LR}_i$ the estimate ultimate loss ratio, obtained by the product of r_i^* and $\hat{m}^*$. These two last elements represent the loss ratio index and the selected individual loss ratio. For the details on the procedure refer to [14].

Regarding the uncertainty in this prior, we choose a strong informative prior such that $CV(\mu_i) = 0$.

For the settlement speed we choose a non-informative prior, with mean given by the estimation obtained by the ODP calibration.

3. BBF2

This version of BBF model is based on the same framework of *BBF1*, but assuming not-informative priors for both μ_i and γ_j. From [8] we already know that the estimation of claims reserve from this model will be close to the ODP Bootstrap one. In fact, using non-informative priors the estimate of claims reserve will be mainly dependent on the underlying ODP model assumption for incremental payments, as in ODP Bootstrap.

One problem that arises in the practical application is related with the autocorrelation of MCMC samples, which is particularly relevant under this model assumptions. The classical solution we adopted in this case is to apply a thinning of the samples, by selecting only a sample of the chain every *n* (where this value depends on the degree of autocorrelation).

4. BBF3

This version of the BBF model is based on the theoretical framework described in Section 2, assuming that the prior information for the ultimate losses is based on the product between earned premiums and market ultimate loss ratio.

In practice, the mean of prior parameter μ_i is obtained as:

$$\mathbb{E}(\mu_i) = B_i \cdot LR_i \tag{4}$$

where LR_i is the expected ultimate loss ratio of the company.

In this case we know the amount of earned premium for each accident year from the dataset (see Table 1), however we do not know the target loss ratio of the company. For this reason we need a proxy to use in place of the company's loss ratio, since we do not have this information. In this case we use the market loss ratio of the US companies for the same LoB. This LR represents the ratio between incurred losses and earned premiums at market level, as reported in NAIC report [23]. Hence, in Formula (4) we substitute LR_i with LR^m, which is the loss ratio at market level, fixed for each accident year.

For the variability of the prior ultimate cost, we choose a strong informativity $CV(\mu_i) = 0$, while for the γ_j prior we use the same assumptions as *BBF1*.

5. BBF4

In this version we choose a market-based approach for both priors of BBF model.

For the ultimate cost prior μ_i we use the same mean and variance as in *BBF3*. For the settlement speed prior we set the mean at market level, computed as the ratio of cumulative payments over incurred losses, calculated on the last diagonal, for the aggregate of companies in the

Table 4

Summary of BBF models prior parameters

	μ_i		γ_j	
Model	**Mean**	**CoV**	**Mean**	**CoV**
BBF1	Mack BF	0	ODP estimate	∞
BBF2	Mack BF	∞	ODP estimate	∞
BBF3	Market LR	0	ODP estimate	∞
BBF4	Market LR	0	Market data	0
BBF5	Market LR	0.5	Market data	0.5

market. More specifically, for a generic development year *j*, the formula for computing the settlement speed is:

$$s_j = \frac{\sum_{c \in C} P^c_{j,n-j}}{\sum_{c \in C} I^c_{j,n-j}} \tag{5}$$

where *C* is the number of companies we have at disposal for that LoB.

Regarding the uncertainty in the γ_j prior we choose a strong informativity, with $CV(\gamma_j) = 0$.

6. BBF5

This last alternative is based on the same calibration of *BB4*, but setting the uncertainty of both the priors as informative, rather than strong informative, with $CV(\mu_i) = CV(\gamma_j) = 0.5$.

We summarize in Table 4 the assumptions of the different BBF models we applied. As reported in the Table the aim is to test the effects on the prediction and on the volatility of the claims reserve of different approaches for the calibrations of the priors μ_i and γ_j providing alternative solutions for the informativity of the priors.

4. Analysis

4.1 *Performance metrics*

In the claims reserving context we are interested in predicting the future development in a way such that our estimate is close to the actual (ex-post) realization. Therefore, in the analysis of predictive performance we compare the estimate of claims reserve obtained from each method with the real outcome. In particular we are interested in measuring the predictive capability of Bayesian Bornhuetter-Ferguson models using as reference benchmark a classic approach, represented by ODP Bootstrap.

In order to perform this task we define some metrics that we use for evaluating the performance of the methods. Having at disposal synthetic indicators, we can perform the analysis using all the simulation, but also looking separately at the performance in each LoB.

In particular, we choose the Mean Error (ME) and Mean Absolute Error (MAE). ME is defined as:

$$\mathrm{ME} = \frac{1}{N}\sum_{c=1}^{N}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{R}^{c,s} - R^{c}}{R^{c}}\right)\right) \tag{6}$$

where $\hat{R}^{c,s}$ is the estimated claims reserve of the model for company *c* and simulation *s* and R^{c} the actual claims reserve realized *ex post* for the same company in the dataset.

For a fixed LoB, the formula is a simple average of the normalized error (with respect to the actual outcome) over the set of N companies in the dataset and over the set of S simulations for each company.

MAE is instead defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{c=1}^{N}\left(\frac{1}{S}\sum_{s=1}^{S}\left|\frac{\hat{R}^{c,s} - R^{c}}{R^{c}}\right|\right) \tag{7}$$

where in this case each error is calculated by means of an absolute value in order to avoid the compensation of positive and negative errors for different simulations and companies.

The sign of ME shows if the model tends to produce an underestimation or overestimation respect to the actual outcome. However, it shall be noted that the ME is subject to possible compensations of positive and negative errors, which could result in a lower aggregate value, still in presence of high errors in opposite directions for different companies. MAE solves this drawback, by means of the absolute value; however it loses the information on the sign of error. Hence, we can have a complete view only considering both indicators.

In addition to the average measures of distances between estimate and outcome, we also assess which model was better able to provide the closest estimate of claims reserve at the valuation date, with respect to the actual (ex post) realization. For this scope, Formula (6) has been also applied separately for each company. Hence, for each company, we rank the models and we count how many times a model showed the best performance.

For insurance companies, in addition to the estimation of claims reserving, it is fundamental to correctly predict also the next year payments. In fact, from an asset-liability management point-of-view it represents the amount that has to be available next year for the payments of the incurred claims.

Actually, insurance companies are interested in computing the expected outflows for each calendar year, and not only the following one. This task is related with the accounting frameworks defined by IFRS and Solvency II, which require the discounting of future expected payments according to the year in which the cash-flow arises.

For this reason we compute the same statistics described above in Formula (6)-(7) for each diagonal of the run-off triangle. However, the latest diagonals are composed of just few elements, which would lead to a high instability of the errors, being based on just one observation. For this reason we split the analysis in just two parts: next year payments and other payments (after next year). In practice, the mean error of next diagonal and other diagonals are computed according to the following formulas:

$$\text{ME}_{\text{ND}} = \frac{1}{N}\sum_{c=1}^{N}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\sum_{i=1}^{n}\left(\hat{P}_{i,n-i+1}^{c,s} - P_{i,n-i+1}^{c}\right)}{\sum_{i=1}^{n} P_{i,n-i+1}^{c}}\right)\right) \tag{8}$$

$$\text{ME}_{\text{O}} = \frac{1}{N}\sum_{c=1}^{N}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\sum_{i+j>n+1}\left(\hat{P}_{i,n-i+1}^{c,s} - P_{i,n-i+1}^{c}\right)}{\sum_{i+j>n+1} P_{i,n-i+1}^{c}}\right)\right) \tag{9}$$

The mean absolute error statistic is obtained by means of the same formulas, replacing the differences in brackets with their absolute values.

Finally, we focus on the uncertainty provided by each model. To this aim, we compute the coefficient of variation (CV) of the results of a LoB (see formula (10)) considering only companies with amount of earned premiums B^c between defined lower and upper boundaries (l and u).

$$CV(\hat{R}^{c,s} \mid (l < B^c <= u)) = \frac{\sigma(\hat{R}^{c,s} \mid (l < B^c <= u))}{E(\hat{R}^{c,s} \mid (l < B^c <= u))} \tag{10}$$

For the choice of lower and upper bound it is possible to use different approaches. The general logic is to split companies in groups whose behavior can be considered homogeneous for their size. In practice, a simple way is to divide between small, medium and large size companies. However, the difficult point is to establish the boundaries for which a companies can be considered in one of this groups.

The other problem is that we could come up with groups having unbalanced number of companies. For this reason we chose to divide in groups of the same size, based on the amount of earned premiums. More specifically, we have computed the quartiles of the earned premium and have divided the companies in four groups based on the quartile they belong to.

4.2 *Predictive analysis*

Accident year

In this section we perform a predictive analysis of claims reserving models, separately for each line of business. We start considering the commercial auto LoB. In Table 5 we report the summary of the metrics described in the previous section for each of the six models that have been tested.

Looking at the ME we can observe that all the models tend to produce on aggregate an overestimation of claims reserve respect to the actual outcome. In particular, Bootstrap produces the lowest error, followed by BBF5; while, according to the MAE statistic, BBF5 shows the lowest error.

In the table we have also reported 4 quantiles representing respectively 10%, 25%, 75% and 90% of the ME distribution. They are particularly useful for assessing the spread of the error distribution, with also information of the tendency to under/over estimation, for each model.

Regarding the values of this 4 quantiles of error distribution, BBF4 has the closest values to 0 for the down side, while Bootstrap and BBF5 for the upper side. In general, we can observe that all the models tend to have $q_{0.1}$ and $q_{0.25}$ closest to 0 respect to $q_{0.75}$ and $q_{0.9}$. This result is in line with the findings that the models for this LoB produce a general overestimation. Finally, we can observe that BBF5 produced the highest number of closest prediction with respect to the observed outcome.

This result shows that, for this dataset, BBF models with priors based on market data are able to produce an improvement in the estimation of claims reserve with respect to other priors' setting and the ODP Bootstrap. In addition, the use of mid-informative priors produces a better result respect to high/non informative ones, since this setting takes better into account both exogenous and endogenous information.

Table 5

Commercial Auto: predictive performance (values in percentage). ME and MAE represent respectively the mean error and the mean absolute error between the estimated claims reserve of each model and the actual outcome. Q10, Q25, Q75 and Q90 represents the quantiles of the respective order, computed on the ME distribution. N* represents the number of closest estimate to the actual outcome for each model. Best performance of each indicator is reported in bold.

	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
ME	**17.80**	58.96	65.95	63.37	42.16	26.30
MAE	44.08	76.48	85.59	82.04	54.73	**39.05**
Q10	-45.35	-32.99	-39.20	-35.91	**-23.78**	-24.63
Q25	-18.93	-12.27	-12.33	-11.64	**-5.55**	-5.55
Q75	**35.60**	92.52	106.92	96.97	73.73	41.78
Q90	85.20	214.22	232.05	235.93	123.36	**82.39**
N*	6	12	3	3	11	**15**

Table 6 describes the result for the private auto line of business. Also for this LoB, the global tendency of all the models is to overestimate the claims reserve. BBF5 results the model with the best predictive performance according to both ME and MAE.

Table 6

Private Auto: predictive performance (values in percentage).

ME and MAE represent respectively the mean error and the mean absolute error between the estimated claims reserve of each model and the actual outcome. Q10, Q25, Q75 and Q90 represents the quantiles of the respective order, computed on the ME distribution. N* represents the number of closest estimate to the actual outcome for each model. Best performance of each indicator is reported in bold.

	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
ME	21.73	59.44	63.65	67.55	15.76	**20.89**
MAE	28.42	65.13	69.48	72.85	30.43	**27.12**
Q10	-13.09	-10.36	-10.85	**-7.38**	-26.74	-12.33
Q25	**1.48**	3.04	4.25	11.08	-9.04	1.60
Q75	34.84	41.17	49.34	49.21	**28.97**	34.20
Q90	**58.53**	233.11	233.79	238.97	74.18	61.53
N*	5	10	1	4	**24**	6

Regarding quantiles of error distribution, there is no single model showing a closest range.

Finally, looking at the model with highest number of closest prediction we observe that BBF4 was the best one for 24 companies over 50, which is almost 50% of the cases. Hence, from this result we can conclude that on average BBF5 produced the lowest error, but BBF4 had the highest number of closest prediction.

This result confirms the findings commented for the commercial auto LoB of improvement in model's performance by using market data in the priors.

Results for workers' compensation line of business are reported in Table 7. In this case the best performance is obtained by the ODP Bootstrap in terms of both ME and MAE. Looking at the quantiles for the BBF models it is possible to notice that they produced a higher overestimation of claims reserve respect to ODP Bootstrap. In particular this is evident for BBF4 and BBF5, which means that market loss ratio was generally higher than the actual loss ratio of the companies.

However, for a peculiar line of business, like worker's compensation, relying on market data can be a solution not optimal. In fact, differently from more traditional businesses where there is a competition between a high number of players which leads to similar performances, the outcome tends to be more company-specific.

Table 7

Workers' compensation: predictive performance (values in percentage).

ME and MAE represent respectively the mean error and the mean absolute error between the estimated claims reserve of each model and the actual outcome. Q10, Q25, Q75 and Q90 represents the quantiles of the respective order, computed on the ME distribution. N* represents the number of closest estimate to the actual outcome for each model. Best performance of each indicator is reported in bold.

	ODP Bootstrap	**BBF1**	**BBF2**	**BBF3**	**BBF4**	**BBF5**
ME	**19.42**	28.77	28.82	42.64	74.09	29.57
MAE	**35.91**	39.72	41.32	52.85	79.62	38.13
Q10	-29.21	-22.07	-24.43	-20.27	**4.76**	-13.60
Q25	-8.32	-5.31	-6.28	2.59	27.39	**0.08**
Q75	**38.77**	44.11	44.31	67.39	101.24	45.63
Q90	**81.45**	93.47	96.56	117.31	174.01	84.04
N*	**11**	10	5	4	5	10

Table 8

Other liability: predictive performance (values in percentage).

ME and MAE represent respectively the mean error and the mean absolute error between the estimated claims reserve of each model and the actual outcome. Q10, Q25, Q75 and Q90 represents the quantiles of the respective order, computed on the ME distribution. N* represents the number of closest estimate to the actual outcome for each model. Best performance of each indicator is reported in bold.

	ODP Bootstrap	**BBF1**	**BBF2**	**BBF3**	**BBF4**	**BBF5**
ME	**2.28**	38.02	744.70	41.36	49.49	19.54
MAE	114.20	57.40	768.92	63.72	71.55	**44.80**
Q10	-60.36	**-35.82**	-46.36	-41.64	-43.83	-46.40
Q25	-31.67	**-12.41**	-17.51	-16.20	-19.53	-21.75
Q75	53.49	61.47	85.76	61.09	92.60	**49.76**
Q90	127.14	137.53	187.89	148.55	157.54	**93.06**
N*	9	9	3	**10**	5	9

In Table 8 we display the results for other liabilities line of business. In this case, the application of ODP Bootstrap produced the lowest prediction error according to ME, however this result is mainly due to a compensation of under and overestimations observed for different companies. In fact, considering the absolute errors, this methodology gives one of the worse results. Looking at the quantiles of error distribution, BBF1 and BBF5 shows the smallest ranges; while for the highest number of best prediction there is not a complete predominance from a single model.

Calendar year analysis

We focus now on the predictions in a calendar year view. Considering Commercial auto (see Table 9), results are in line with the accident year analysis, with the BBF5 model assuring the best performance.

It is noticeable for the private auto LoB the different prediction performance between next year payments and payments of the following years (see Table 10). Indeed, BBF4 and BBF5 models show the best performance in predicting the next-year payments, according respectively to ME and MAE. For the following years, instead, ODP Bootstrap shows the best behavior.

In Table 11 is reported the calendar year analysis for the workers' compensation LoB. Differently from what we observed in a accident year

Table 9

Commercial Auto: predictive performance - calendar year view (values in percentage)

	ODP Bootstrap	**BBF1**	**BBF2**	**BBF3**	**BBF4**	**BBF5**
ME_{ND}	**37.78**	44.37	54.33	51.30	63.30	40.22
ME_{O}	**22.23**	98.94	106.37	105.12	55.73	36.98
MAE_{ND}	61.23	60.48	71.58	68.81	77.55	**57.71**
MAE_{O}	58.34	125.30	134.42	130.51	73.15	**54.74**

Table 10

Private Auto: predictive performance - calendar year view (values in percentage)

	ODP Bootstrap	**BBF1**	**BBF2**	**BBF3**	**BBF4**	**BBF5**
ME_{ND}	15.45	19.64	22.68	24.72	**10.95**	15.12
ME_{O}	**39.70**	129.61	136.42	143.32	40.91	41.42
MAE_{ND}	22.22	25.32	28.33	29.84	23.31	**21.90**
MAE_{O}	**50.57**	138.95	145.82	151.52	61.19	51.21

Table 11

Workers' compensation: predictive performance - calendar year view (values in percentage)

	ODP Bootstrap	**BBF1**	**BBF2**	**BBF3**	**BBF4**	**BBF5**
ME_{ND}	46.35	**45.76**	50.48	57.95	96.26	52.02
ME_{O}	**32.05**	58.36	58.74	75.90	118.56	55.08
MAE_{ND}	59.65	**56.01**	61.76	67.20	100.78	61.88
MAE_{O}	**54.73**	75.13	76.95	90.04	125.54	66.37

framework, the BBF1 appears as the best model for predicting next-year payments. Bootstrap-ODP remains the best choice in the long run.

Considering other liability LoB, we observed in Table 12, that 4 out of 6 models shows a prevalence of underestimation in the prediction of next year payments. However, this behavior is mainly due to the presence of some outliers in the claims development for specific companies.

Regarding the predictive performance, also in this case ODP Bootstrap shows the best results according to ME, while BBF5 for MAE.

Table 12

Other liability: predictive performance - calendar year view (values in percentage)

	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
ME_{ND}	**-0.23**	-6.48	2194.12	-10.41	29.59	-9.80
ME_{O}	**-1.52**	58.52	660.91	71.35	59.98	29.69
MAE_{ND}	137.09	104.54	2304.10	99.39	123.42	**92.34**
MAE_{O}	139.56	83.49	690.52	99.81	90.13	**60.88**

In addition, looking at MAE, we can observe that all the models had a poor performance in predicting next-year payments. For instance, the BB5 model, who has the best performance, shows an error of 92.34%, which means that, on average, the insurance company should be asked to pay around the double (or the half) of what expected from the model. This result shows the difficulties of using a fixed scheme, represented by the blind application of a model without any additional information, on a peculiar LoB as other liability.

4.3 *Volatility analysis*

In this section we analyze the effect of the size of a company on the variability of the estimation of claims reserve for the models analyzed. Since we are analyzing companies of different size we chose a relative measure, consisting in the coefficient of variation of claims reserve estimate, in order to measure the uncertainty.

The theoretical expectation is that a large company shall have a lower CV respect to one of small size. It means that the relative variability in the estimate of claims reserve decreases, increasing the size of the company. We are interested in assessing if this theoretical expectation is also realized empirically on the selected dataset.

In Figure 1 the coefficient of variation obtained from the application of ODP Bootstrap and BBF5 on the commercial auto dataset are compared. We have chosen ODP Bootstrap since it is the usual reference model for this analysis and BBF5 as a representative of Bayesian models since it was the one with the best predictive performance in most of the cases. We can observe that both models show a decreasing trend for the coefficient of variation for increasing volume. It is also noticeable how on average the bootstrap model leads to a higher uncertainty.

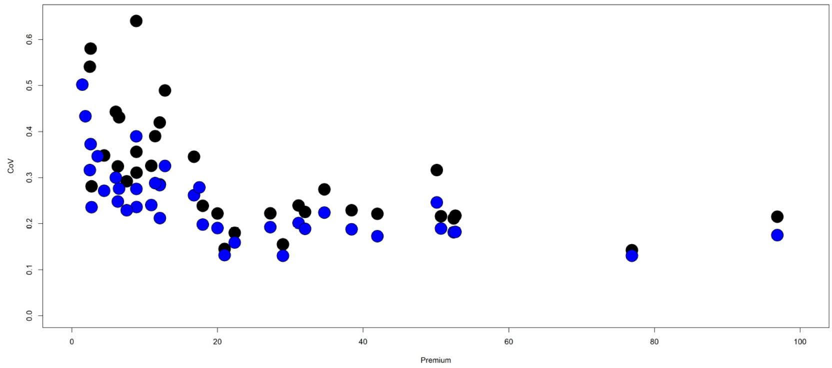

Figure 1

Commercial Auto: pattern of coefficient of variations of claims reserve according to volume of premiums. In black ODP Bootstrap, in blue BBF5. On the x-axis, we report the earned premiums of the specific company, while on the y-axis we have the CV of the models. The range size of earned premium is in thousand of USD

In order to have detailed numerical results, we have applied Formula (10) for each model. In particular, we have divided the companies in four groups according to the amount of earned premiums as described in Section 4.1.

The results of this analysis are reported in Table 13 and show that when the size range increases the CV decreases. BBF4 model shows the lowest CV, since it assumes a strong informativity for both its priors leading to small standard deviation of estimated claims reserve.

Table 13
Commercial Auto: volatility analysis
(values in percentage)

Size range	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
1,432 – 8,866	246.11	35.65	47.78	34.66	21.75	32.27
8,866 – 24,812	98.17	31.33	38.32	28.81	13.62	23.37
24,812 – 91,889	22.24	23.78	28.14	23.69	11.17	18.54
91,889 – 3,543,796	12.50	21.36	22.71	21.88	6.09	11.16

The same trend can be found also for the other lines of business, with few different patterns, especially in the other liability LoB, probably related to outliers and the small sample size.

Table 14

Private Auto: volatility analysis (values in percentage)

Size range	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
3,810 – 53,575	19.57	24.23	27.13	23.88	11.06	17.24
53,575 – 134,272	10.32	17.59	19.33	17.28	6.55	9.73
134,272 – 288,238	8.97	31.19	32.08	30.42	5.55	8.47
288,238 – 117,655,840	7.31	17.78	18.50	17.40	4.40	6.96

Table 15

Workers' compensation: volatility analysis (values in percentage)

Size range	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
2,898 – 38,753	193.06	30.56	35.14	30.34	12.64	21.48
38,753 – 115,901	15.03	20.73	23.54	19.31	6.17	12.56
115,901 – 310,557	12.32	14.02	16.11	13.34	4.69	10.82
310,557 – 2,905,415	8.98	6.94	8.95	7.62	3.58	8.03

Table 16

Other liability: volatility analysis (values in percentage)

Size range	ODP Bootstrap	BBF1	BBF2	BBF3	BBF4	BBF5
486 – 4,043	1867.05	27.56	80.80	27.09	20.80	31.10
4,043 – 15,730	901.33	29.46	1973.55	28.18	15.67	27.27
15,730 – 44,522	62.37	27.48	36.76	27.67	11.61	22.17
44,522 – 2,414,413	1452.56	17.89	31.78	18.06	7.04	16.19

5. Conclusion

The importance of a correct estimation of claims reserve has become a central objective for insurance companies, because of the effects in their balance sheet and in the assessment of risk. In addition, the necessity of having a measure of uncertainty around this estimate lead to the proposal of several stochastic models. In this paper we have focused on the analysis of the performance of Bayesian Bornhuetter-Ferguson (BBF) model. The reason of this choice is that this model gives the possibility of using different sources of information and appropriately weight its priors in order to estimate the claims reserve. We see this approach as one that should be pursued by insurance companies, which can rely on their expertise and knowledge of their portfolio in order to calibrate the prior parameters and update them as they obtain additional external information.

We have analyzed five alternative models based on the Bayesian Bornhuetter-Ferguson model, using as reference benchmark the traditional ODP Bootstrap, on a set of American companies belonging to four different LoBs. According to the predictive metrics, the best performance has been obtained by the BBF5 for most of the indicators and LoBs analyzed. This specific model is based on market data for both its (informative) priors. Also for the calendar year analysis this model has provided in most of the cases the closest estimate respect to the actual realization, among the BBF-based models. In the comparison with ODP Bootstrap the performance was strongly dependent on the chosen metric, with the former having the best performance with ME and the latter with MAE.

As in general for Bayesian models, their strength of using external information for the calibration of their priors can also became their weakness in case these information proves inaccurate. On this dataset, however, we have shown that just using (informative) global market data proves enough to obtain accurate prediction in-line with ODP Bootstrap and in some cases even better.

References

[1] Alai, D. H., Merz, M., and Wüthrich, M. V. (2011). Prediction uncertainty in the Bornhuetter-Ferguson claims reserving method: revisited. *Annals of Actuarial Science*, 5(1) : 717.

[2] Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. In *Proceedings of the casualty actuarial society*, volume 59, pages 181-195.

[3] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1) : 1-26.

[4] Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall.

[5] EIOPA (2015). Guidelines on the valuation of technical provisions. Technical report, EIOPA.

[6] England, P. (2002). Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving". *Insurance: Mathematics and Economics*, 31(3) : 461-466.

[7] England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3) : 443-518.

[8] England, P. D., Verrall, R. J., and Wüthrich, M. V. (2012). Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method. *Annals of Actuarial Science*, 6(2) : 258-283.

[9] Friedland, J. F. (2009). *Estimating Unpaid Claims Using Basic Techniques*. CAS.

[10] Gao, G. (2016). *Three essays on Bayesian claims reserving methods in general insurance*. PhD thesis, College of Business and Economics.

[11] Gilks, W. R., Best, N. G., and Tan, K. (1995). Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(4) : 455-472.

[12] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) : 97-109.

[13] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, 23(2) : 213-225.

[14] Mack, T. (2006). Parameter Estimation for Bornhuetter/Ferguson. *Casualty Actuarial Society Forum*, pages 141-157.

[15] Mack, T. (2008). The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin: The Journal of the IAA*, 38(1) : 87-103.

[16] Martinek, L. (2019). Analysis of stochastic reserving models by means of NAIC claims data. *Risks*, 7(2) : 62.

[17] Martinek, L., Arato, M., and Malyusz, M. (2015). Comparison of stochastic claims reserving models in insurance. Available on ArXiv.

[18] Martínez-Miranda, M., Nielsen, J., and Verrall., R. (2013). Double chain ladder and bornhuetter-ferguson. *North American Actuarial Journal*, 17 : 101-113.

[19] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6) : 1087-1092.

[20] Meyers, G. (2015). *Stochastic loss reserving using Bayesian MCMC models*. Casualty Actuarial Society.

[21] Meyers, G. and Shi, P. (2011a). Loss Reserving Data Pulled From NAIC Schedule P. Available at http://www.casact.org/research/index.cfm.

[22] Meyers, G. G. and Shi, P. (2011b). The retrospective testing of stochastic loss reserve models. In Casualty Actuarial Society E-Forum, Summer.

[23] NAIC (2004). Market Share Reports for property/casualty groups and companies.Technical report, National Association of Insurance Commissioners. https://www.naic.org/prod_serv/MSR-PB-05.pdf.

[24] NAIC (2020). Uniform property & casualty product coding matrix. Technical report, National Association of Insurance Commissioners. https://www.naic.org/documents/industry_pcm_p_c_2020.pdf.

[25] Pinheiro, P. J., Andrade e Silva, J. M., and de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4) : 701-714.

[26] Renshaw, A. and Verrall, R. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(04) : 903-923.

[27] Shapland, M. R. (2016). *Using the ODP Bootstrap Model: A Practitioner's Guide*. Casualty Actuarial Society.

[28] Shi, P., Basu, S., and Meyers, G. G. (2012). A bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, 16 : 29-51.

[29] Shi, P. and Frees, E. W. (2011). Dependent loss reserving using copulas. *ASTIN Bulletin*, 41 : 449-486.

[30] Tee, L. and Käärik, M. and Vii, R. (2017). On comparison of stochastic reserving methods with bootstrapping. *Risks*, 5(2) : 1-21.

[31] Wang, S. and Chen, H. (2012). Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data. *Variance*, 8(2) : 182-203.

[32] Wüthrich, M. V. (2017). Non-life insurance: mathematics & statistics. Available at SSRN.

[33] Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

[34] Wüthrich, M. V. and Merz, M. (2015). Stochastic claims reserving manual: Advances in dynamic modeling. Technical Report 15-34, Swiss Finance Institute Research Paper.

[35] Wüthrich, M. V. (2010). Accounting year effects modeling in the stochastic chain ladder reserving method. *North American Actuarial Journal*, 14(23) : 5-55.

Received April, 2021

Revised July, 2021