Prediction Uncertainty in the Bornhuetter-Ferguson Claims Reserving Method

Daniel Alai Michael Merz Mario Wüthrich

ETH Zurich

August 1, 2009

Overview

- Data and notation.
- Model considerations.
- The Bornhuetter-Ferguson predictor.
- Maximum likelihood estimation of the model parameters.
- Prediction uncertainty.
- Numerical example.

The Data

- Let $X_{i,j}$ denote the incremental claims of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, I\}$.

- At time I, we have observations $\mathcal{D}_I = \{X_{i,j},\ i + j \leq I\}$.

- We predict the corresponding lower triangle $\{X_{i,j},\ i + j > I\}$.

- Define $C_{i,j}$ to be the cumulative claims of accident year i up to development year j,

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}.$$

The Data

- Let $X_{i,j}$ denote the incremental claims of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, I\}$.
- At time I, we have observations $\mathcal{D}_I = \{X_{i,j},\ i + j \leq I\}$.
- We predict the corresponding lower triangle $\{X_{i,j},\ i + j > I\}$.
- Define $C_{i,j}$ to be the cumulative claims of accident year i up to development year j,

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}.$$

The Data

- Let $X_{i,j}$ denote the incremental claims of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, I\}$.
- At time I, we have observations $\mathcal{D}_I = \{X_{i,j},\ i + j \leq I\}$.
- We predict the corresponding lower triangle $\{X_{i,j},\ i + j > I\}$.
- Define $C_{i,j}$ to be the cumulative claims of accident year i up to development year j,

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}.$$

The Data

- Let $X_{i,j}$ denote the incremental claims of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, I\}$.
- At time I, we have observations $\mathcal{D}_I = \{X_{i,j},\ i + j \leq I\}$.
- We predict the corresponding lower triangle $\{X_{i,j},\ i + j > I\}$.
- Define $C_{i,j}$ to be the cumulative claims of accident year i up to development year j,

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}.$$

A Visual Representation

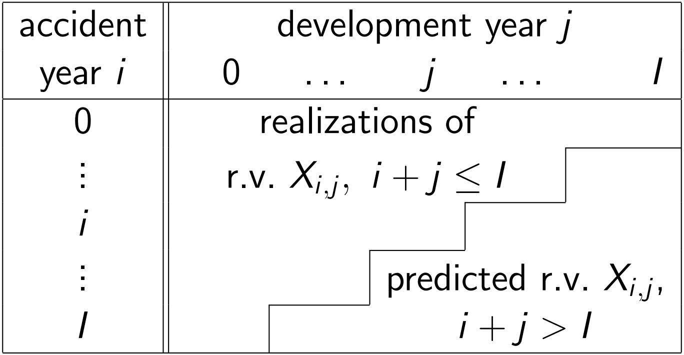

Figure: Claims development triangle.

Model Assumptions (ODP)

- The incremental claims $X_{i,j}$ are independent overdispersed Poisson distributed (ODP) with

$$\begin{aligned} E\left[X_{i,j}\right] &= m_{i,j} = \mu_i \gamma_j, \\ \mathrm{Var}\left(X_{i,j}\right) &= \phi\, m_{i,j}, \end{aligned}$$

  and $\sum_{j=0}^{I} \gamma_j = 1$.

- $\hat{\nu}_k$ are independent unbiased estimators of the expected ultimate claim $\mu_k = E\left[C_{k,I}\right]$ for all $k \in \{0, \ldots, I\}$.
- $X_{i,j}$ and $\hat{\nu}_k$ are independent for all i, j, k.

Remark: For MSEP considerations, an estimate of the uncertainty of the $\hat{\nu}_k$ is required. We assume that a prior variance estimate $\widehat{\mathrm{Var}}(\hat{\nu}_i)$ is given exogenously.

Model Assumptions (ODP)

- The incremental claims $X_{i,j}$ are independent overdispersed Poisson distributed (ODP) with

$$
\begin{aligned}
E\left[X_{i,j}\right] &= m_{i,j} = \mu_i \gamma_j, \\
\mathrm{Var}\left(X_{i,j}\right) &= \phi\, m_{i,j},
\end{aligned}
$$

  and $\sum_{j=0}^{I} \gamma_j = 1$.
- $\hat{\nu}_k$ are independent unbiased estimators of the expected ultimate claim $\mu_k = E\left[C_{k,I}\right]$ for all $k \in \{0, \ldots, I\}$.
- $X_{i,j}$ and $\hat{\nu}_k$ are independent for all i, j, k.

Remark: For MSEP considerations, an estimate of the uncertainty of the $\hat{\nu}_k$ is required. We assume that a prior variance estimate $\widehat{\mathrm{Var}}(\hat{\nu}_i)$ is given exogenously.

Model Assumptions (ODP)

- The incremental claims $X_{i,j}$ are independent overdispersed Poisson distributed (ODP) with

$$\begin{aligned} E\left[X_{i,j}\right] &= m_{i,j} = \mu_i \gamma_j, \\ \mathrm{Var}\left(X_{i,j}\right) &= \phi\, m_{i,j}, \end{aligned}$$

  and $\sum_{j=0}^{I} \gamma_j = 1$.
- $\hat{\nu}_k$ are independent unbiased estimators of the expected ultimate claim $\mu_k = E\left[C_{k,I}\right]$ for all $k \in \{0, \ldots, I\}$.
- $X_{i,j}$ and $\hat{\nu}_k$ are independent for all i, j, k.

Remark: For MSEP considerations, an estimate of the uncertainty of the $\hat{\nu}_k$ is required. We assume that a prior variance estimate $\widehat{\mathrm{Var}}(\hat{\nu}_i)$ is given exogenously.

Model Assumptions (ODP)

- The incremental claims $X_{i,j}$ are independent overdispersed Poisson distributed (ODP) with

$$\begin{aligned} E\left[X_{i,j}\right] &= m_{i,j} = \mu_i \gamma_j, \\ \mathrm{Var}\left(X_{i,j}\right) &= \phi\, m_{i,j}, \end{aligned}$$

  and $\sum_{j=0}^{I} \gamma_j = 1$.
- $\hat{\nu}_k$ are independent unbiased estimators of the expected ultimate claim $\mu_k = E\left[C_{k,I}\right]$ for all $k \in \{0, \dots, I\}$.
- $X_{i,j}$ and $\hat{\nu}_k$ are independent for all i, j, k.

Remark: For MSEP considerations, an estimate of the uncertainty of the $\hat{\nu}_k$ is required. We assume that a prior variance estimate $\widehat{\mathrm{Var}}(\hat{\nu}_i)$ is given exogenously.

The Bornhuetter-Ferguson Predictor

- In practice, the Bornhuetter-Ferguson (BF) predictor relies on the data $\mathcal{D}_I$ for the loss development pattern and on external data or expert opinion for the expected ultimate claims $E[C_{i,I}]$.

- The ultimate claim $C_{i,I}$ of accident year i is predicted by

$$\widehat{C}_{i,I}^{BF} = C_{i,I-i} + \hat{\nu}_i \sum_{j>I-i} \hat{\gamma}_j,$$

where $\hat{\gamma}_j$ is an estimator for γ_j.

Maximum Likelihood Estimators for ODP

$$
\begin{aligned}
l_{\mathcal{D}_I}(\mu_i, \gamma_j, \phi) &= \sum_{\substack{i+j \le I \\ j < I}} \left(\frac{1}{\phi} (X_{i,j} \log(\mu_i \gamma_j) - \mu_i \gamma_j) + \log c(X_{i,j}; \phi) \right) \\
&+ \left(\frac{1}{\phi} (X_{0,I} \log \left[\mu_0 \left(1 - \sum_{n=0}^{I-1} \gamma_n \right) \right] - \mu_0 \left(1 - \sum_{n=0}^{I-1} \gamma_n \right) + \log c(X_{0,I}; \phi) \right),
\end{aligned}
$$

where $c(\cdot, \phi)$ is the suitable normalizing function.

The development pattern obtained, $\hat{\gamma}_j$, is identical to that produced by the chain ladder method,

$$\hat{\gamma}_j = \hat{\gamma}_j^{CL}.$$

Maximum Likelihood Estimators for ODP

$$
\begin{aligned}
l_{\mathcal{D}_I}(\mu_i, \gamma_j, \phi) &= \sum_{\substack{i+j\leq I \\ j<I}} \left(\frac{1}{\phi}(X_{i,j} \log(\mu_i \gamma_j) - \mu_i \gamma_j) + \log c(X_{i,j}; \phi) \right) \\
&+ \left(\frac{1}{\phi}(X_{0,I} \log \left[\mu_0 \left(1 - \sum_{n=0}^{I-1} \gamma_n \right) \right] - \mu_0 \left(1 - \sum_{n=0}^{I-1} \gamma_n \right) + \log c(X_{0,I}; \phi) \right),
\end{aligned}
$$

where $c(\cdot, \phi)$ is the suitable normalizing function.

The development pattern obtained, $\hat{\gamma}_j$, is identical to that produced by the chain ladder method,

$$\hat{\gamma}_j = \hat{\gamma}_j^{CL}.$$

Dispersion Parameter Estimation

- To estimate the dispersion parameter ϕ, one could use MLE.
- We do not. Instead, due to ease of implementation we use Pearson residuals, given by

$$\hat{\phi} = \frac{1}{d} \sum_{i+j \leq I} \frac{(X_{i,j} - \hat{m}_{i,j})^2}{\hat{m}_{i,j}},$$

where $d = \frac{(I+1)(I+2)}{2} - 2I - 1$ is the degrees of freedom of the model and $\hat{m}_{i,j} = \hat{\mu}_i \hat{\gamma}_j$.

Dispersion Parameter Estimation

- To estimate the dispersion parameter ϕ, one could use MLE.
- We do not. Instead, due to ease of implementation we use Pearson residuals, given by

$$\hat{\phi} = \frac{1}{d} \sum_{i+j \leq I} \frac{(X_{i,j} - \hat{m}_{i,j})^2}{\hat{m}_{i,j}},$$

where $d = \frac{(I+1)(I+2)}{2} - 2I - 1$ is the degrees of freedom of the model and $\hat{m}_{i,j} = \hat{\mu}_i \hat{\gamma}_j$.

Mean Square Error of Prediction

The (conditional) mean square error of prediction (MSEP) of the BF predictor $\widehat{C}_{i,I}^{BF}$ for single accident years $i \in \{1, \ldots, I\}$ is given by

$$\mathrm{msep}_{C_{i,I}|\mathcal{D}_I}(\widehat{C}_{i,I}^{BF}) = E\left[\left(\widehat{C}_{i,I}^{BF} - C_{i,I}\right)^2 \middle| \mathcal{D}_I\right]$$

$$= \sum_{j>I-i} \mathrm{Var}(X_{i,j}) + \left(\sum_{j>I-i} \hat{\gamma}_j\right)^2 \mathrm{Var}(\hat{\nu}_i) + \mu_i^2 \left(\sum_{j>I-i} \hat{\gamma}_j - \sum_{j>I-i} \gamma_j\right)^2.$$

We treat the three terms separately.

Development Pattern Uncertainty

We estimate

$$\Big(\sum_{j>I-i}(\hat{\gamma}_j-\gamma_j)\Big)^2$$

by the unconditional expectation

$$E\Bigg[\Big(\sum_{j>I-i}(\hat{\gamma}_j-\gamma_j)\Big)^2\Bigg]=\sum_{\substack{j>I-i\\k>I-i}}E\Big[\Big(\hat{\gamma}_j-\gamma_j\Big)\Big(\hat{\gamma}_k-\gamma_k\Big)\Big].$$

Neglecting that MLEs have a possible bias term we make the following approximation:

$$\sum_{\substack{j>I-i\\k>I-i}}E\Big[\Big(\hat{\gamma}_j-\gamma_j\Big)\Big(\hat{\gamma}_k-\gamma_k\Big)\Big]\approx\sum_{\substack{j>I-i\\k>I-i}}\mathrm{Cov}(\hat{\gamma}_j,\hat{\gamma}_k).$$

Development Pattern Uncertainty

We estimate

$$\Big(\sum_{j>I-i}(\hat{\gamma}_j-\gamma_j)\Big)^2$$

by the unconditional expectation

$$E\left[\Big(\sum_{j>I-i}(\hat{\gamma}_j-\gamma_j)\Big)^2\right]=\sum_{\substack{j>I-i\\k>I-i}}E\left[\Big(\hat{\gamma}_j-\gamma_j\Big)\Big(\hat{\gamma}_k-\gamma_k\Big)\right].$$

Neglecting that MLEs have a possible bias term we make the following approximation:

$$\sum_{\substack{j>I-i\\k>I-i}}E\left[\Big(\hat{\gamma}_j-\gamma_j\Big)\Big(\hat{\gamma}_k-\gamma_k\Big)\right]\approx\sum_{\substack{j>I-i\\k>I-i}}\mathrm{Cov}(\hat{\gamma}_j,\hat{\gamma}_k).$$

Asymptotic Properties of the MLE

In order to quantify the parameter estimation uncertainty $\hat{\gamma}_j - \gamma_j$ we use the asymptotic MLE property

$$\sqrt{n}\Big(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\Big) \stackrel{(d)}{\rightarrow} \mathcal{N}\Big(\mathbf{0}, H(\boldsymbol{\zeta}, \phi)^{-1}\Big), \text{ as } n \rightarrow \infty,$$

with Fisher information matrix $H(\boldsymbol{\zeta}, \phi) = (h_{r,s}(\boldsymbol{\zeta}, \phi))_{r,s=1,\ldots,m}$, given by

$$h_{r,s} = -E_{\boldsymbol{\zeta}}\left[\frac{\partial^2}{\partial\zeta_r \partial\zeta_s} l_{\mathcal{D}_I}(\boldsymbol{\zeta}, \phi)\right],$$

for $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{2I+1}) = (\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_{I-1})$ and $\hat{\boldsymbol{\zeta}}$ the corresponding MLE.

By replacing the parameters in $h_{r,s}$ by their estimates, we obtain $\hat{h}_{r,s}$.

Asymptotic Properties of the MLE

In order to quantify the parameter estimation uncertainty $\hat{\gamma}_j - \gamma_j$ we use the asymptotic MLE property

$$\sqrt{n}\left(\hat{\zeta} - \zeta\right) \overset{(d)}{\rightarrow} \mathcal{N}\left(\mathbf{0}, H(\zeta, \phi)^{-1}\right), \text{ as } n \rightarrow \infty,$$

with Fisher information matrix $H(\zeta, \phi) = (h_{r,s}(\zeta, \phi))_{r,s=1,\dots,m}$, given by

$$h_{r,s} = -E_{\zeta}\left[\frac{\partial^2}{\partial \zeta_r \partial \zeta_s} l_{\mathcal{D}_I}(\zeta, \phi)\right],$$

for $\zeta = (\zeta_1, \dots, \zeta_{2I+1}) = (\mu_0, \dots, \mu_I, \gamma_0, \dots, \gamma_{I-1})$ and $\hat{\zeta}$ the corresponding MLE.

By replacing the parameters in $h_{r,s}$ by their estimates, we obtain $\hat{h}_{r,s}$.

Estimate of the MSEP

The estimator for the (conditional) MSEP for a single accident year $i \in \{1, \ldots, I\}$ is given by

$$\widehat{\mathrm{msep}}_{C_{i,I}|\mathcal{D}_I}(\widehat{C}_{i,I}^{BF}) = \sum_{j>I-i} \hat{\phi}\, \hat{\nu}_i \hat{\gamma}_j + \left(\sum_{j>I-i} \hat{\gamma}_j \right)^2 \widehat{\mathrm{Var}}(\hat{\nu}_i) + \hat{\nu}_i^2 \sum_{\substack{j>I-i \\ k>I-i}} \hat{h}_{j,k}.$$

The estimator for the (conditional) MSEP for aggregated accident years is given by

$$\widehat{\mathrm{msep}}_{\sum_{i=1}^I C_{i,I}|\mathcal{D}_I}\left(\sum_{i=1}^{I} \widehat{C}_{i,I}^{BF} \right) = \sum_{i=1}^{I} \widehat{\mathrm{msep}}_{C_{i,I}|\mathcal{D}_I}(\widehat{C}_{i,I}^{BF}) + 2 \sum_{i<n} \hat{\nu}_i \hat{\nu}_n \sum_{\substack{j>I-i \\ k>I-n}} \hat{h}_{j,k}.$$

Numerical Example

We analyze the dataset given below, given in 000's.

i/j	0	1	2	3	4	5	6	7	8	9	$\hat{\nu}_i$
0	5,947	3,721	896	208	207	62	66	15	11	16	11,653
1	6,347	3,246	723	152	68	37	53	11	12		11,367
2	6,269	2,976	847	263	153	65	54	9			10,963
3	5,863	2,683	723	191	133	88	43				10,617
4	5,779	2,745	654	274	230	105					11,045
5	6,185	2,828	573	245	105						11,481
6	5,600	2,893	563	226							11,414
7	5,288	2,440	528								11,127
8	5,291	2,358									10,987
9	5,676										11,618

Table: Observed incremental payments $X_{i,j}$ and prior estimates $\hat{\nu}_i$.

Furthermore, we assume the uncertainty of the $\hat{\nu}_i$ to be given by a coefficient of variation of 5%. Hence,

$$\widehat{\mathrm{Var}}(\hat{\nu}_i) = \hat{\nu}_i^2 \, (0.05)^2.$$

Results for AMW Method (2009)

acc. year i	BF reserves	process std. dev.	prior std. dev.	parameter std. dev.	prior and parameter std. dev.	msep$^{1/2}$	Vco
1	16,120	15,401	806	15,539	15,560	21,893	135.8%
2	26,998	19,931	1,350	17,573	17,624	26,606	98.5%
3	37,575	23,514	1,879	18,545	18,639	30,005	79.9%
4	95,434	37,473	4,772	24,168	24,635	44,845	47.0%
5	178,023	51,181	8,901	29,600	30,910	59,790	33.6%
6	341,305	70,866	17,065	35,750	39,614	81,187	23.8%
7	574,089	91,909	28,704	41,221	50,231	104,739	18.2%
8	1,318,645	139,294	65,932	53,175	84,703	163,025	12.4%
9	4,768,385	264,882	238,419	75,853	250,195	364,362	7.6%
cov.				195,409	195,409	195,409	
total	7,356,575	329,007	249,828	228,249	338,396	471,971	6.4%

Table: Reserve and uncertainty results for single and aggregated accident years.

Comparison to Mack Method (2008) and the Chain Ladder Method

	reserves	process error	estimation error	$\text{msep}^{1/2}$	Vco
BF AMW (2009)	7,356,575	329,007	338,396	471,971	6.4%
BF Mack (2008)	7,505,506	621,899	375,424	726,431	9.7%
CL method	6,047,061	424,379	185,026	462,960	7.7%

Table: Aggregate reserve and uncertainty results for the CL method, the BF approach of A., Merz, Wüthrich, and the BF approach of Mack (2008).

Mack (2008) utilizes a different development pattern.

Comparison to Mack Method (2008) and the Chain Ladder Method

	reserves	process error	estimation error	$\text{msep}^{1/2}$	Vco
BF AMW (2009)	7,356,575	329,007	338,396	471,971	6.4%
BF Mack (2008)	7,505,506	621,899	375,424	726,431	9.7%
CL method	6,047,061	424,379	185,026	462,960	7.7%

Table: Aggregate reserve and uncertainty results for the CL method, the BF approach of A., Merz, Wüthrich, and the BF approach of Mack (2008).

Mack (2008) utilizes a different development pattern.

Cumulative Development Pattern Uncertainty

j	AMW (2009)		Mack (2008)	
	$\hat{\beta}_j$	s.e.$(\hat{\beta}_j)$	$\hat{z}_j^*$	s.e.$(\hat{z}_j^*)$
0	58.96%	0.653%	58.60%	1.717%
1	88.00%	0.484%	87.66%	0.616%
2	94.84%	0.370%	94.60%	0.326%
3	97.01%	0.313%	96.84%	0.271%
4	98.45%	0.258%	98.35%	0.131%
5	99.14%	0.219%	99.07%	0.054%
6	99.65%	0.175%	99.62%	0.025%
7	99.75%	0.160%	99.73%	0.018%
8	99.86%	0.137%	99.85%	0.012%
9	100.00%		100.00%	

Table: Cumulative development pattern, a comparison.

References

Alai, D. H., Merz, M., & Wüthrich, M. V. (2009). Mean square error of prediction in the Bornhuetter-Ferguson claims reserving method. To appear in Annals of Actuarial Science.

Bornhuetter, R. L. & Ferguson, R. E. (1972). The actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, **LIX**, 181–195.

Mack, T. (2008). The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin*, **38**, 87–103.

Verrall, R. J. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, **8**, 67–89.

Thank you!