JAB Chain: A model-based calculation of paid and incurred loss development factors

Bertrand Verdier
Swiss Reinsurance Company
Mythenquai 50/60, P.O. Box, CH-8022 Zürich, Switzerland
bertrand_verdier@swissre.com
tel: +41 43 285 63 67

Artur Klinger
Alea Europe
Kornhausgasse 7, P.O. Box, CH-4002, Basel, Switzerland
artur.klinger@aleagroup.com
tel: +41 61 268 22 11

Abstract

The analysis of reinsurance treaties often requires a separate analysis of both paid and incurred claims. However, such a method can lead to very different ultimate projections. Quarg and Mack (2003) recently proposed a solution to this well known problem in his presentation of Munich Chain Ladder. They proposed then to take into account the previous paid amounts together with the paid to incurred ratios (i.e. the quotient of paid and incurred amounts). Whereas they used an iterative procedure on the normalized residuals, here we set up a linear mixed model, which allows us to get the full development in a single step and to use the related theoretical results, eg for standard error estimation.

Keywords

Time-varying effect of paid to incurred ratios, state-space models (Kalman filtering), penalized least-squares, smoothing, JAB Chain, Munich Chain Ladder.

Acknowledgements

This thesis was done during my final actuarial internship at Alea Europe, under the supervision of Prof. Partrat, from the Institut de Sciences Financières et d'Assurance (ISFA).

Introduction

For estimation of ultimate losses cumulated paid losses and incurred losses are usually developped independently. Although one would expect that both methods finally come to a similar result, this is not always the case in practice. One reason is that interdependecies between both processes may exist which are not taken into account by independently estimating paid and incurred loss development factors.

Quarg and Mack presented in 2003 Munich Chain Ladder [7], we will refer to by MCL in the following. This method was specially designed to remedy the drawback mentioned above. This method requires past figures for both paid and incurred as usual, but also the paid to incurred ratios. The key idea is using the reserving information included in the incurred amounts to improve for the paid losses. In particular for recent accident years (AY) years, there are few development years so the paid amounts are usually low: the major part of the information is in the incurred position. For example, if the incurred amount is relatively high, that means that the reserving department expects relatively high paid losses for the following years. This information can be incorporated by comparing the paid to incurred ratios for each business year to the average ratio.

Without presenting the method in its entirety, we would like to comment on how the concept for the JAB Chain developed from its predecessor:

- time varying slopes: as stated in MCL presentation, the correlation coefficients λ^P and λ^I, ie the slopes used in the model are assumed to be constant over DY. However, as Quarg and Mack noted, this is not the case in practice. Considering all the DY together is equivalent to assume that the influence of a P/I ratio variation is equal for all development years. But the information carried by the incurred losses is sensible for the paid claims development in the early DY whereas after several years, the claims just develop "normally", and taking into account the incurred process is less informative.

- integration into one model: MCL uses the reserving information included in the incurred process to model the corresponding paid amounts. Nevertheless, the paid process is also used to calculate the incurred losses. In our model, we want to separate this. While the incurred process may be informative to the paid losses, it is not necessary that the paid process gives additional information on the incurred. The incurred amounts are supposed to be determined by using all historical payment information. Therefore, the incurred process had to be modelled separately from the paid process. The first one might be modelled differently e.g. by incorporating also calendar year effects.

- joint estimation of all factors: MCL lies on regression residual analysis, and thus implies estimating some parameters, calculating the residuals, then deducing the slopes to finally predict the future paid and incurred amounts. Instead, we adopted a single model, which allows to estimate simultaneously all parameters and thus takes into account possible dependencies between them. This also allows direct calculations for model validation and standard errors. This can be found in the theoretical literature about semiparametric and mixed models (see Ruppert, Wand and Carroll (2003) [8]).

Chapter 1

JAB Chain

According to what was exposed previously, we introduce a general model-based framework assuming that the incurred losses include additional information which can be used for the paid claims development.

1.1 Model presentation

1.1.1 Assumptions

According to the previous comments, our assumptions are the following ones:

- The cumulated paid amount $P_{i,j}$ depends on the previous paid amount $P_{i,j-1}$ and the previous incurred amount, or to be accurate, the previous P/I ratio $Q_{i,j-1}$.
- The incurred amount $I_{i,j}$ only depends on the previous incurred amount $I_{i,j-1}$.

where i=2..n denotes the accident year (AY) and j=2...n denotes the development year (DY).

The first assumption illustrates the link between paid and incurred process we have been dealing with since the beginning. As we explained in the previous comments, the second is different from the MCL assumptions: we only use the incurred information to model the paid amounts. The incurred process is assumed to keep all relevant information on historical claims.

Rather than using several residuals analysis, we chose to use a single and direct model. It belongs to the linear state space models family, a detailed presentation of which can be found in Fahrmeir and Tutz (2001) [3]. The key idea is modeling the paid development factors and taking into account the relative level of the P/I ratio, ie in comparison to its average q_j over all accident years. Thus, depending on the variation to the average, a correction is applied to the reference development factor.

Since we assume that the influence of the P/I is not the same over the different DY, a coefficient is introduced, varying over DY.

Let us now write the *linear observation equation*, which gives the general form of the model:

$$P_{i,j+1} = P_{i,j} \cdot (\alpha_j + \beta_j \cdot (Q_{i,j} - q_j)) + \epsilon_{i,j} \tag{1.1}$$

$$\boxed{\frac{P_{i,j+1}}{P_{i,j}} = \alpha_j + \beta_j \cdot (Q_{i,j} - q_j) + \frac{\epsilon_{i,j}}{P_{i,j}} \qquad \text{where} \quad \begin{cases} i, j = 1 \dots n-1 \\ \{\epsilon_{i,j}\}_{j=1 \dots n-1} \sim \left[0, \sigma_j^{P2} \cdot P_{i,j}\right] \end{cases}}$$

$Q_{i,j}$ is the corresponding P/I ratio. Those equations show explicitely the model heteroscedasticity: the $\epsilon_{i,j}$ variance is proportional to the payments $P_{i,j}$.

Please note that in contrast to MCL, we only use the P/I ratios, which can easily be interpreted. Indeed, the P/I ratio is the share of the global claim (paid and reserves) which has been already paid, and it is generally between 0 and 1, whereas the I/P vary between 1 and infinity (c.f. Quarg and Mack [7]).

The equation above can be derived from the usual CL model equation, by adding a second term. This $\beta_j \cdot (Q_{i,j} - q_j)$ corresponds to the correction due to the variation of the P/I ratio to its average, which is assumed to vary over DY. In the CL model, α_j is the development factor f_j.

The linear state space model environment allows us to assume smoothingly varying α_j and β_j parameters. The sequence of states is defined by two *linear transition equations*:

$$\boxed{\begin{aligned} \alpha_{j+1} = &\quad \alpha_j + \epsilon_{\alpha_j} \\ \beta_{j+1} = &\quad \beta_j + \epsilon_{\beta_j} \end{aligned}} \tag{1.2}$$

where α_j and β_j are so called *random walks*. ϵ_{α_j} and ϵ_{β_j} are white noise processes (noted WNP), such that: $\mathbf{E}(\alpha_j) = \mathbf{E}(\beta_j) = 0$, $\mathbf{Var}\left(\epsilon_{\alpha_j}\right) = \sigma_\alpha^2$, $\mathbf{Var}\left(\epsilon_{\beta_j}\right) = \sigma_\beta^2$.

To illustrate the way the smoothing works, let us show some key values for the varying parameters σ_α and σ_β. For example, $\sigma_\alpha \to 0$ implies that α_j gets constant to α_{j-1}. Thus, all the α_j get equal to the same constant. On the other hand $\sigma_\alpha \to \infty$, ie assuming an infinite variance, is equivalent to say that α_j can take any value with the same probability. In this case, the α_j are not smoothed and the fit to the observed is optimal. The graphs presented on fig. 1.1 show this evolution: the β_j parameters are constant for $\sigma_\beta \to 0$ and their volatility increase as σ_β tends to infinity.

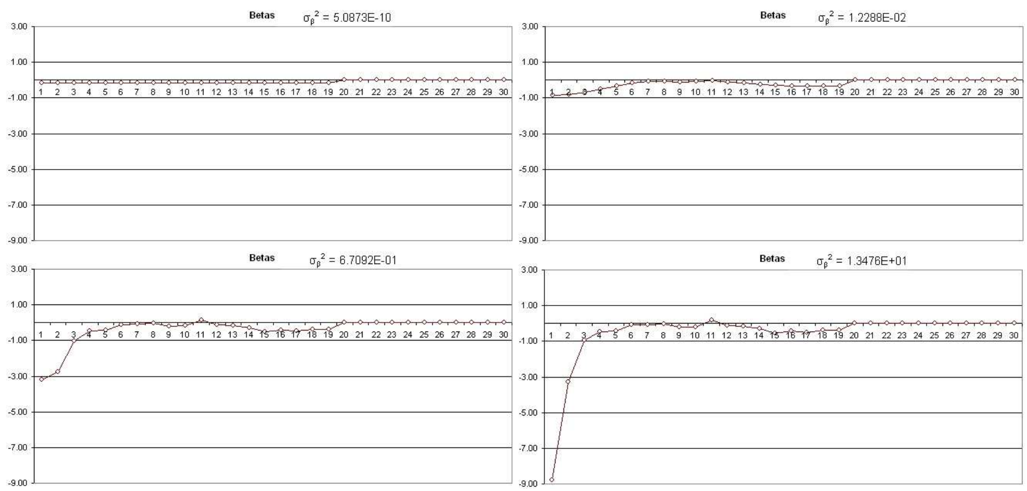


Figure 1.1: β_j estimation depending on σ_β

Fig. 1.2 shows the same evolution for the α_j coefficients. Please note that in addition to that, the β_j coefficients were estimated as σ_α was kept constant equal to $1.26E-05$ and the α_j were estimated as σ_β was equal to $5.09E-10$.

1.1.2 Parameters estimation: α_j, β_j

The parameters estimation $(\alpha_j)_{j=1\dots n-1}$ and $(\beta_j)_{j=1\dots n-1}$ is done following a *weighted penalized least square* procedure. The penalty derived from the linear transition equations. Indeed, let us assume gaussian errors for the transition equations, this means that α_j and β_j are modelled by:

$$\begin{aligned} \epsilon_{\alpha_j} &\sim N(0, \sigma_\alpha) \\ \epsilon_{\beta_j} &\sim N(0, \sigma_\beta) \end{aligned} \qquad \forall j = 1 \dots n-1$$

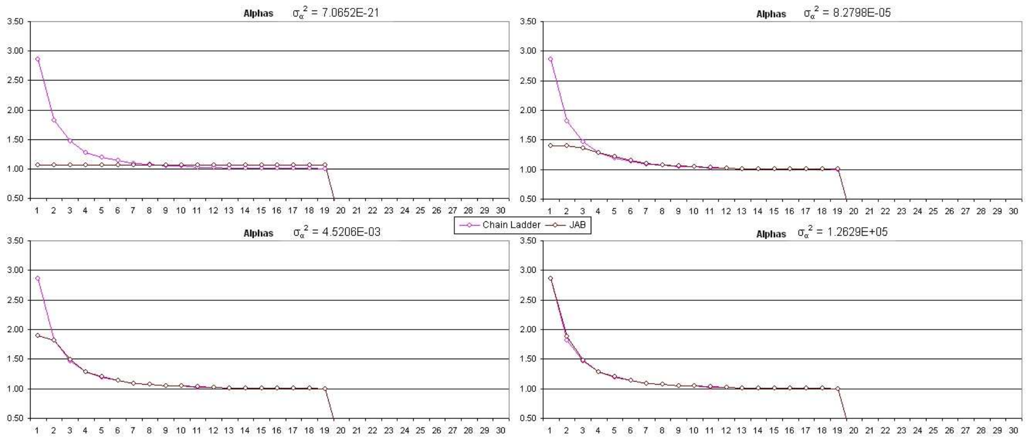


Figure 1.2: α_j estimation depending on σ_α

Under these assumptions, the related likelihood is:

$$L(\alpha_1, \ldots, \alpha_{n-1}) = \prod_{j=1}^{n-2} \frac{1}{\sigma_\alpha \cdot \sqrt{2\pi}}\, e^{\frac{-\left(\alpha_{j+1}-\alpha_j\right)^2}{2\sigma_\alpha^2}}$$

And the log-likelihood corresponds to:

$$\ln\left(L(\alpha_1, \ldots, \alpha_{n-1})\right) = -\frac{1}{2\sigma_\alpha^2} \cdot \sum_{j=1}^{n-2} (\alpha_{j+1} - \alpha_j)^2 - C \qquad \text{where} \quad C = (n-2)\ln\left(\sigma_\alpha \sqrt{2\pi}\right)$$

Note there are no constraints on α_1 and β_1. This can be understood as a diffuse prior distribution on the starting values, i.e. no assumptions about the level of the coefficients are incorporated into the model.
Lastly, as there is not enough data to estimate two paramaters on top-right edge of the triangle, i.e α_{n-1} and β_{n-1} are not identifiable. Therefore we impose the restriction $\beta_{n-1} = 0$, i.e the effect of the P/I ratio tends to zero for the last DY.

Consequently, according to the linear transition equations and to the gaussian assumption for the errors, the penalty of the fit criterion can be written as:

$$\left\{\begin{array}{ll} \min\limits_{\substack{\alpha_j \, / \\ j=2,\dots,n-1}} & \dfrac{1}{\sigma_\alpha^2} \cdot \sum_{j=1}^{n-2} \left(\alpha_{j+1} - \alpha_j\right)^2 \\ \min\limits_{\substack{\beta_j \, / \\ j=2,\dots,n-2}} & \dfrac{1}{\sigma_\beta^2} \cdot \left(\sum_{j=1}^{n-3} \left(\beta_{j+1} - \beta_j\right)^2 + \beta_{n-2}^2 \right) \end{array}\right. \tag{1.3}$$

And:

$\beta_{n-1} = 0$

No explicit condition on α_1 and β_1

The equations 1.3 give the conditions deriving from the linear transition equations (1.2). To get the whole criterion leading to the final estimators for the varying coefficients α_j and β_j, you must add the least-square criterion. Then $(\alpha_j)_{j=1\dots n}$ and $(\beta_j)_{j=1\dots n}$ minimize the quantity, so called M (penalized least square criterion):

$$M = \sum_{i=1}^{n} M_i \qquad \text{where}$$

$$M_i = \sum_{j=1}^{n-1} \omega_{i,j} \left(P_{i,j+1} - \beta_j \left(Q_{i,j} - q_j\right) P_{i,j} - \alpha_j P_{i,j}\right)^2 + \frac{1}{\sigma_\alpha^2} \sum_{j=1}^{n-2} \left(\alpha_{j+1} - \alpha_j\right)^2 + \frac{1}{\sigma_\beta^2} \left(\sum_{j=1}^{n-3} \left(\beta_{j+1} - \beta_j\right)^2 + \beta_{n-2}^2 \right) \tag{1.4}$$

This form of smoother allows, via σ_α and σ_β, a trade-off between goodness of fit and smoothness. Indeed, the first part of the equation is the least-square term, and the two last terms penalize high variation in the α_j and β_j. This way of finding those corresponding parameters is called penalized least-square estimation.

Parameters estimation: the weights $w_{i,j}$

The weighted least square criterion (1.4) includes weights depending on DY and AY. Indeed, the variance is generally not constant: the equation 1.1 shows that the conditional variance given the history of payments up to DY j:

$$\mathbf{Var}\left(P_{i,j+1} \mid \mathbf{P}_i(j)\right) = \sigma_j^{P^2} \cdot P_{i,j}$$

which is consistent with the definition of $\sigma_j^{P^2}$ given in the MCL presentation:

$$\mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}} \mid \mathbf{P}_i(j)\right) = \frac{\sigma_j^{P^2}}{P_{i,j}}$$

If we drop the β_j term in our model, we get the CL model:

$$P_{i,j+1} = P_{i,j} \cdot f_j + \epsilon_{i,j}$$

Thus,

$$\epsilon_{i,j} = P_{i,j+1} - P_{i,j} \cdot f_j$$

This may be estimated by:

$$\widehat{\epsilon_{i,j}} = P_{i,j+1} - P_{i,j} \cdot \widehat{f_j}$$

On the other hand, our model assumption gives:

$$\mathbf{Var}\left(\epsilon_{i,j} \mid \mathbf{P}_i(j)\right) = P_{i,j} \cdot \sigma_j^{P^2}$$

which yields to:

$$\mathbf{Var}\left(\frac{\epsilon_{i,j}}{\sqrt{P_{i,j}}} \mid \mathbf{P}_i(j)\right) = \sigma_j^{P^2}$$

$$\mathbf{Var}\left(\frac{P_{i,j+1} - P_{i,j} \cdot f_j}{\sqrt{P_{i,j}}} \mid \mathbf{P}_i(j)\right) = \sigma_j^{P^2}$$

thus, $\sigma_j^{P^2}$ can be estimated as usually by the estimator:

$$\widehat{\sigma_j^{P^2}} = \widehat{\mathbf{Var}}\left(\sqrt{P_{i,j}} \cdot \left(\frac{P_{i,j+1}}{P_{i,j}} - \hat{f}_j^P\right) \mid \mathbf{P}_i(j)\right)$$

$$\widehat{\sigma_j^{P^2}} = \frac{1}{n-j-1} \cdot \sum_{i=1}^{n-j} P_{i,j} \cdot \left(\frac{P_{i,j+1}}{P_{i,j}} - \hat{f}_j^P\right)^2$$

in complete analogy to Quarg and Mack. Since the model residuals are scaled by the previous years payments, we obtain the weights in the penalized least squares criterion as:

$$w_{i,j} = \frac{1}{\widehat{\mathbf{Var}}\left(\epsilon_{i,j} \mid \mathbf{P}_i(j)\right)} = \frac{1}{\sigma_j^{P^2} \cdot P_{i,j}}$$

Implemention: the switch to the matrix form

In order to simplify the implementation, it is worth to rewrite the equations 1.4 in a matrix manner. Let γ stands for the column-vector of α_j and β_j parameters and $\hat{\gamma}$ its estimate. With appropriately defined matrix A_i, B, C_i, K and W_i (diagonal matrix of weights ω_j), one can get:

$$\boxed{M = \sum_{i=1}^{n} (BC_i - A_i\gamma)^t W_i (BC_i - A_i\gamma) + \gamma^t K \gamma} \tag{1.5}$$

Finally, if $\sum_{i=1}^{n} A_i^t W_i A_i + K$ is non singular:

$$\boxed{\hat{\gamma} = \left(\sum_{i=1}^{n} A_i^t W_i A_i + K \right)^{-1} \cdot \sum_{i=1}^{n} A_i^t W_i B C_i} \tag{1.6}$$

, which can be solved by any matrix package.

1.1.3 Parameters estimation: σ_α, σ_β

As shown on the figures 1.1 and 1.2, the two variance parameters σ_α and σ_β have the role of smoothing parameters and they have to be selected. In many cases the smoothing parameter may be chosen according to visual inspection of appropriate plots. In order two select smoothing parameters automatically we propose following two criterions:

Ultimate P/I ratios optimization

The first criterion consists in finding the variance parameters which lead to the best ultimate P/I ratios. Let us explain what we mean by "best". Indeed, each AY has its own paid and incurred development and and its own P/I ratio, consequently we evaluate the "global quality" of the ultimate P/I ratios by taking a weighted sum of the individual ultimate P/I ratios minus one. The weight for the AY i is $n+1-i$ where n is the number of DY we have. This weighting puts higher importance on old AY where we the forecasting period is shorter. Thus, we try to minimize the following criterion:

$$\boxed{C_{Ult\,PI} = \sum_{i=1}^{n} (n+1-i) \cdot \left(\frac{P_{i,n}}{I_{i,n}} - 1 \right)^2}$$

Generalized cross-validation criterion

The second criterion we implemented is the generalized cross-validation criterion, a description of which is given by Ruppert, Wand and Carroll (2003) [8].

1.2 Results

1.2.1 Data

We present the results of the JAB chain and compare them to the CL and the MCL calculation. The datasets used in the current study come from three different European liability books. In order to anonymize the data, we extracted some losses and multiplied the data by a factor.We will denote them by portfolio no.1, no.2 and no.3

1.2.2 α_j, β_j estimation

Let us now have a look to the estimation of the varying coefficients α_j and β_j coefficients (see fig. 1.3).

We can see that the slopes are negative with falling absolut value, especially for the first portfolio. Contrary to the MCL calculation, we work with the P/I ratios, consequently it is not possible to compare the slopes directly.

As expected, the influence of the incurred process is larger for the first DY. This is confirmed by higher β_j for small j (in absolut values). Then they usually get closer to zero. This can be clearly seen for the the first portfolio, and remain true for the other ones.

1.2.3 Ultimate paid losses, ultimate P/I ratios

The figure 1.4 shows the projected ultimate paid losses and ultimate P/I ratios according to the three methods CL, MCL and JAB Chain. As it can be seen on the

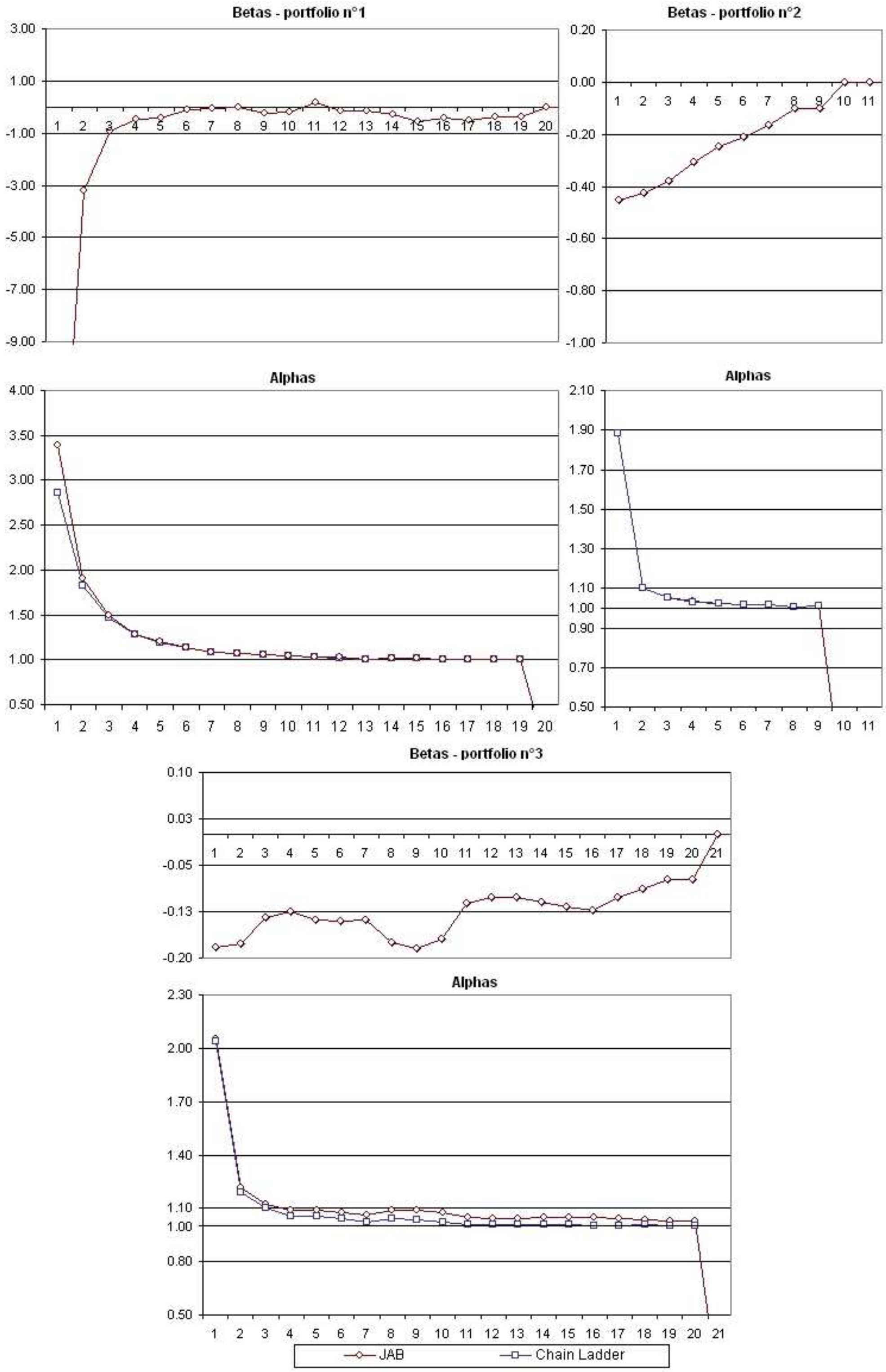


Figure 1.3: Alphas and betas, for the three different portfolios

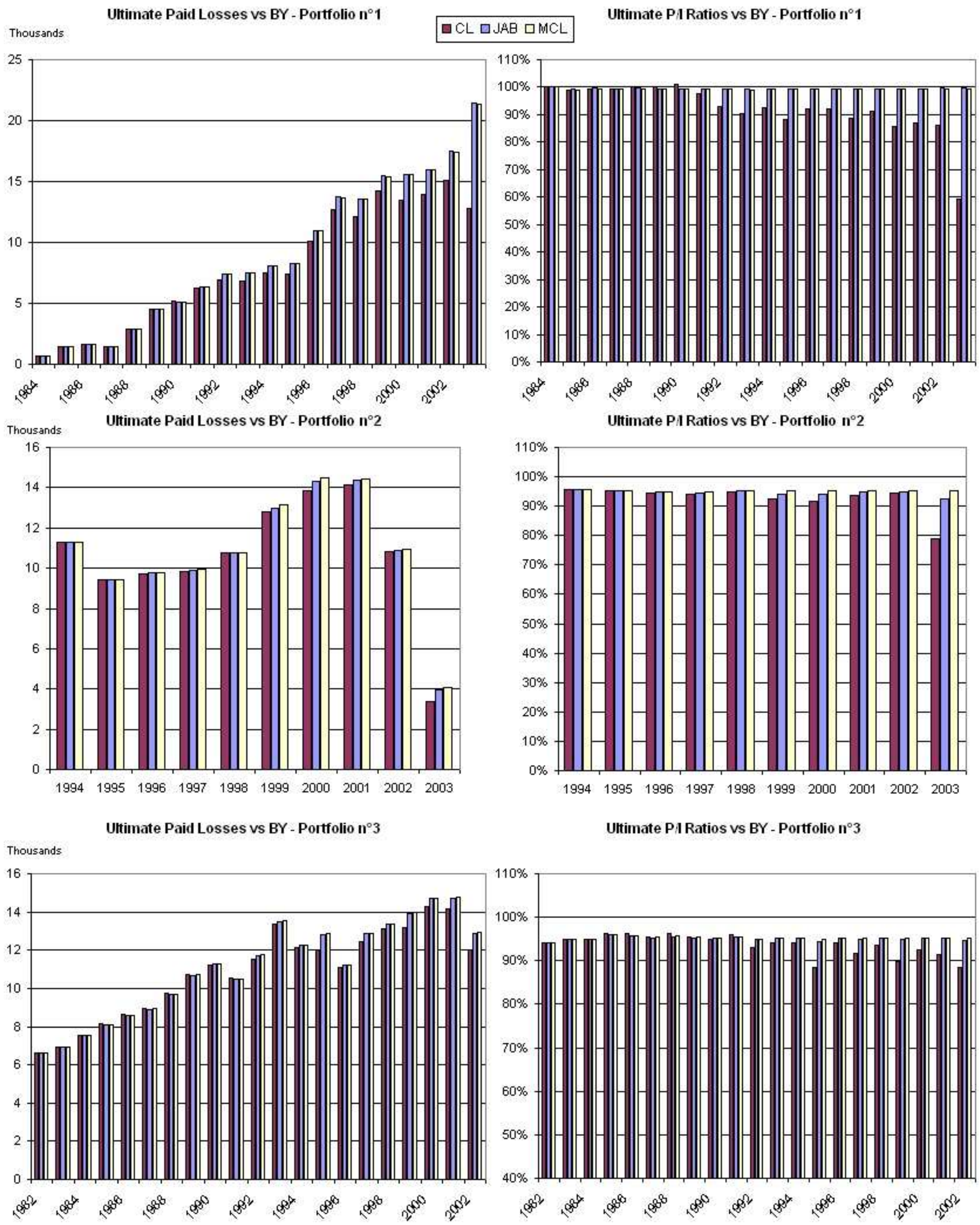


Figure 1.4: Ultimate paid claims and ultimate P/I ratios

P/I ratios plots, both MCL and JAB provide much better results than CL. This indicates that the incurred and the paid development is coherent. If we check the ultimate paids graphs, we see that both MCL and JAB forecast greater ultimate paid claims.

The last point concerns the ultimate P/I ratio: it appears on the graph that for the two last portfolios, the P/I ratio does not converge towards 1 but a little bit below, around 95%. This comes from the historical upper right ratios, which are 94% and 95% for the resp. portfolios no. 2 and 3. Without contradicting our expectations, this shows that both MCL and JAB develop the claims according to the past, which means in other words that if the claims related to the first AY are not fully developed, then the methods "assume" that the other AY will develop in the same period until the same ultimate level. This can be treated by adding a tail factor at the end of the last DY, ie making assumptions for the tail development (eg exponential decay...). On the other hand, it shows a classical drawback of the methods: the ultimate result is very sensitive to the last data in the upper right corner of the triangle.

1.2.4 Patterns

The fig. 1.5 shows the cumulated paid development factors vs DY, for four different AY, 1996, 2000, 2002 and 2003, for the portfolio no. 1. This figure allows us to compare both the ultimate values given by the three methods and the development pattern.

We see again that both MCL and JAB lead to higher ultimate projected paid claims than CL, and the difference is as large as the development period to forecast is long. Furthermore, MCL and JAB forecast more or less the same ultimate values. But the interesting point is that those figures allow a comparison of the different patterns. For old AY, ie when few DY have to be predicted, JAB and MCL give very similar results. On the other hand, for recent AC, ie when many DY have to be predicted, the patterns are different. Indeed, JAB gives higher predictions for the early DY and MCL progressively adjusts it to finally reach the same ultimate level.

This illustrates our previous comment: the correction brought by MCL according to

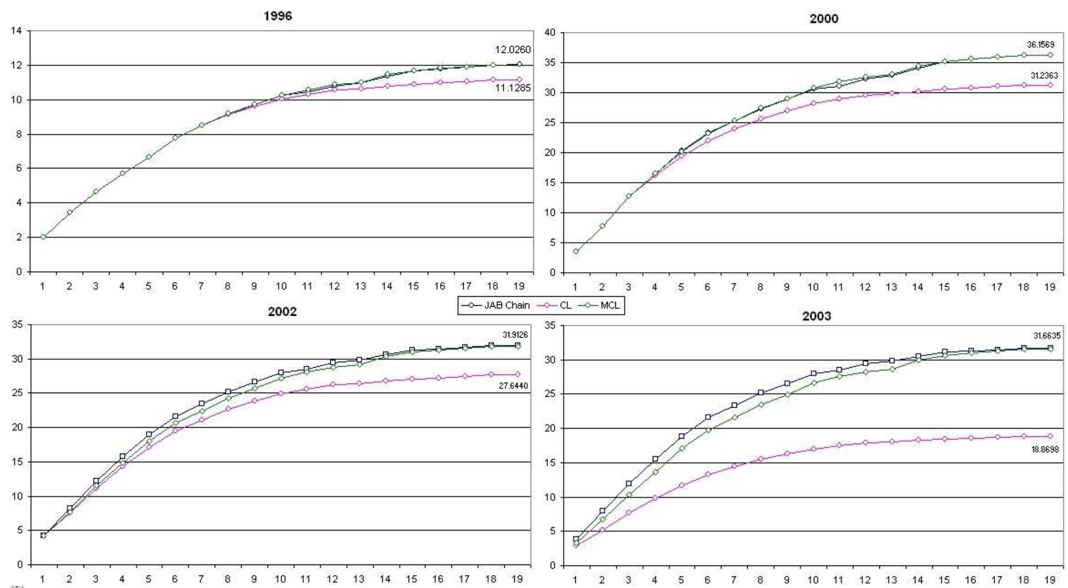


Figure 1.5: Cumulated PDF for several AY

the relative level of P/I ratio is spread over the whole period, because of the constant slope. On the contrary, MCL uses variable coefficients, and the estimations show that they are higher in absolut value for recent early DY. This is in line with our expectations: the reserving process is more informative at the beginning since the paid amounts are still low. Thus, the correction is greater in the early DY for JAB than for MCL but in the end of the development, the two method give more or less the same value.

1.2.5 Validation of model prediction

In order to compare the accuracy of the different methods, we perform the following test: we remove some historical data and estimate the parameters for this information. Then we compare the forecasted values with the historical observations previously removed. Specifically, we remove a certain number of calendar years (ie bottom diagonals on the historical triangle) and then estimate all the parameters. Thus, it is possible to develop the triangle as before and compare the forecasts to the removed observations. The first results were very encouraging, showing undeniably a better goodness of fit for MCL and JAB than for CL, with a slight advantage of

JAB over MCL. Nevertheless, our study was only based on three portfolios and therefore our method would need further tests

Chapter 2

Conclusion and Outlook

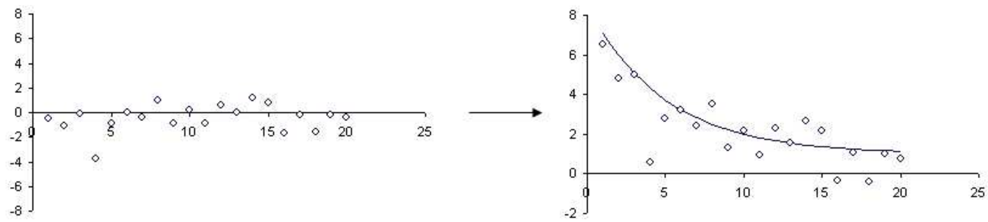
Our model is based on the great idea of MCL, which has been taken over to a uniform modeling framework. This allowed us in particular to assume coefficients varying over time. The results are promising and the framework can simply be adopted to extend MCL in various other directions.

We propose thus a single and global model, based on linear mixed models, which allows us to use the theoretical background, and avoid a several-steps analysis by using a straightforward approach. Moreover we choose to use the incurred information to model the paid process but not the reverse because in our view, the paid process is not informative for the incurred process. Lastly, we only use the incurred information under the form P/I ratio, and allow variable correction, depending on the DY.

Nevertheless, the key idea is very recent and the field remains largely unexplored. Here are some potential developments, among many others:

- Incurred developments factors can be incorporated and estimated simultaneously in one model.
- The model can simply be extended to non-gaussian paid and incurred processes in analogy to Fahrmeir and Tutz (2001). This is of particular value when the distribution of the ultimates should be explored.
- Instead of assuming the coefficients to vary over (DY), the coefficient β_j may also be assumed to vary over a different time scales or even over other metric variables.

- Our model assumes that $\alpha_{t+1} = \alpha_t + \epsilon_t$. This assumption may be changed for a more sophisticated one. For example: $\alpha_{t+1} = a\ \alpha_t + \epsilon_t$. This would lead to



an exponential model, which could be relevant for the $(\alpha_j)_{j=1\ldots n}$ coefficients.

Bibliography

[1] ENGLAND P. D., VERRALL R. J. (2002). *Stochastic claims reserving in general insurance*, Institute of Actuaries.

[2] FAHRMEIR, L., GIEGER C. AND KLINGER A. (1995). *Additive, dynamic and multiplicative regression*, Institüt für Statistik, Universität München.

[3] FAHRMEIR L. AND TUTZ, G. (2001). *Multivariate statistical modeling based on generalized linear models*, 2nd ed, Springer Series in Statistics.

[4] HUTCHINSON, M. F. AND DE HOOG, F. R. (1985). Smoothing noisy data with spline functions. *Numerische mathematik 47*, 99-106.

[5] LIPPE S., SEDLMAIR H., WITTING T. (1990). Praxisrelevante Aspekte der Burning-Cost-Kalkulation. *Deutschen ASTIN-Gruppe*. Presented in Munich in April 1990 and in Travemünde in April 1991.

[6] MACK, T. (1997). *Schriftenreihe Angewandte Versicherungsmathematik, Heft 28: Schadenversicherungsmathematik*, Deutsche Gesellschaft für Versicherungsmathematik.

[7] QUARG, G. AND MACK, T. (2003). Munich Chain Ladder, a reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses.

[8] RUPPERT, D., WAND, M. P. AND CAROLL, R. J. (2003) *Semiparametric regression*, Cambridge Series in Statistical and Probabilistic Mathematics.

[9] VERBEEK H. G. (1972) An approach to the analysis of claims experience in motor liability excess of loss insurance. *Astin Bulletin 6*,195-202.