COHERENT INCURRED PAID (CIP) MODELS FOR CLAIMS RESERVING

BY

GILLES DUPIN, EMMANUEL KOENIG, PIERRE LE MOINE, ALAIN MONFORT AND ERIC RATIARISON

ABSTRACT

In this paper, we first propose a statistical model, called the Coherent Incurred Paid model, to predict future claims, using simultaneously the information contained in incurred and paid claims. This model does not assume log-normality of the levels (or normality of the growth rates) and is semi-parametric since it only specifies the first and the second moments; however, in order to evaluate the impact of the normality assumption, we also propose a benchmark Gaussian version of our model. Correlations between growth rates of incurred and paid claims are allowed and the tail development period is estimated. We also provide methods for computing the Claim Development Results and their Values at Risk in the semi-parametric framework. Moreover, we show how to take into account the updating of the estimation in the computation of the Claim Development Results. An application highlights the practical importance of relaxing the normality assumption and of updating the estimation of the parameters.

KEYWORDS

Incurred and paid claims, simultaneous estimation, correlation, semi-parametric approach, tail development, prediction, updating, CDR, VaR.

1. INTRODUCTION

One of the more important problems that non-life insurance companies have to solve is the evaluation of the reserve risk. Such an evaluation necessitates a two-step modelling. The first step requires a method of prediction of the ultimate claims. The second defines and computes a measure of the reserve risk based on these predictions. Let us consider more precisely these two steps.

There exists a large literature dealing with the first step. Most methods are based either on cumulated payments or on incurred losses. However, there are important works proposing models using both sources of information. Halliwell (1997, 2009) and Venter (2008) used a regression approach. Quarg and Mack

Astin Bulletin 48(2), 749-777. doi: 10.1017/asb.2017.36

(2004) introduced the Munich Chain Ladder (MCL) involving a modification of the Chain Ladder development factors based on incurred-paid ratios. Liu and Verrall (2010) developed a bootstrap estimation of the predictive distributions based on the MCL method. Dahms (2008) and Dahms *et al.* (2009) proposed the complementary loss ratio method (CLRM), which is similar to the MCL method but also allows for the study of uncertainty measures like the mean square errors of predictions (MSEP). Posthuma *et al.* (2008) suggested a multivariate model conditioned on equality of the total paid and incurred losses. Merz and Wüthrich (2010) proposed a probabilistic model, the Paid Incurred Chain (PIC) model, combining in a rigorous way the two kinds of information. This work has been followed by several extensions incorporating new features: tail development factors (Merz and Wüthrich, 2013), dependence (Happ and Wüthrich, 2013; Peters *et al.*, 2014) or individual claims (Pigeon *et al.,* 2014).

The second step of the modelling is the definition and the computation of the reserve risk and it has also been extensively studied. Following recommendations of regulatory authorities, the more popular measure is based on the Claim Development Result (CDR) defined as the difference between the prediction of the ultimate claims today and in 1 year time. More precisely, the measure of the reserve risk is the 99.5% quantile of the opposite of the CDR, and it can be viewed as an evaluation of the under-provisioning: this is the so-called Value at Risk (VaR) notion. (see Wüthrich and Merz (2013)). This measure has been used in the context of the CLRM (Dahms *et al.*, 2009) and in the context of the PIC method (Happ *et al.*, 2012).

In the present paper, we consider both steps of the modelling strategy. We propose a statistical method, the Coherent Incurred Paid (CIP) method, using simultaneously information based on paid and incurred claims. This method is semi-parametric in the sense that it specifies parametrically the first- and second-order moments of the variables of interest but otherwise leaves their probability distributions completely free. However, since the (log) normal distributions have been often considered in the literature, we also propose a benchmark parametric approach based on these distributions. More precisely, in the parametric case, we use a conditional approach in the spirit of Posthuma *et al.* (2008); in the semi-parametric case, we only propose parametric specifications of the expectations and of the variance–covariance matrix of the bidimensional vector composed of the rates of increase of the cumulated payments and incurred losses implying the equality of the payments and of the incurred losses at the ultimate development year. The whole distribution is estimated by non-parametric kernel methods. In the parametric case, the estimation method is the Maximum Likelihood (ML) method, whereas in the semi-parametric approach, the parameters are estimated by the Pseudo Maximum Likelihood (PML) method (see Gourieroux *et al.* (1984)). These methods are flexible enough to incorporate correlation between the rates of increase of the cumulated payments and of the incurred losses at various development years, to allow for specifications also depending on the accident year and to estimate the date at which predictions of the ultimate claims based on both kinds of information become equal. Our approach also allows to compute reserve risk measures, in particular, the CDR and their VaR and it is possible to evaluate the impact of the non-Gaussian features of the variables on these measures as well as their sensitivity to the updating of the parameter estimations at the end of the year defining the CDR's. An application on incurred claims and cumulated payments corresponding to a line of business Motor Body Liability-Insurance highlights several points. First, our CIP method is easily implementable. Second, the projected values of the incurred claims and cumulated payments corresponding to the largest observed development years are very different. Third, these values are also very different from the ones provided by the Chain Ladder method, and the CIP method provides a unique ultimate value which is located between the ultimate values of the Chain Ladder method. Fourth, the results of our CIP method may also be significantly different from those of the PIC methods. Fifth, in the computation of the VaRs of the CDRs, it is crucial to take into account the non-Gaussianity of the rates of increase and the updating of the estimations.

The paper is organized as follows. Section 2 carefully describes the contributions of the paper and their connections with the literature. Section 3 describes the Gaussian CIP model, the estimation of its parameters and of the tail development year, the computation of predictions as well as CDR's and their VaR's. Section 4 generalizes these results to a semi-parametric framework in which Gaussianity is no longer assumed. Section 5 proposes an application. Section 6 provides concluding remarks. Proofs and data are gathered in appendices.

2. Contributions of the paper and comparison with the literature

2.1. Statistical models and methods

One of the main objectives of the present paper is to propose a distribution free approach, only based on a parametric specification of the first- and second-order moments of the variables. Moreover, since the number of parameters does not depend on the number of observations, we avoid the "incidental" issue (see Neyman and Scott (1948), Lancaster (2000) and Moreira (2009)) and we can use the results of the PML theory, in particular, we can test the significativity of the parameters. We also propose a benchmark (log) Gaussian version of our approach; in this case, the statistical method is the standard ML approach and the consequences of the (log) Gaussian assumption can be evaluated.

In comparison, the PIC method is a parametric Bayesian approach (except for the variance parameters) based on independent Gaussian rates of increase of payments and incurred losses, the number of parameters increasing with the size of the triangle. In the PIC method, the payments and the incurred losses are not treated symmetrically. In a first version (Merz and Wüthrich, 2010), the payments are listed first and then incurred losses, implying that the variance of the log-payments (resp. log-incurred losses) is an increasing (resp. decreasing)

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

function of the development year, whereas in a second version (Happ and Wüthrich, 2013), it is the opposite situation. In our approach, we avoid this choice and this asymmetric treatment which could provide different results in particular in the non-Gaussian case. Distribution-free methods are proposed by Liu and Verrall (2010) who use a bootstrap approach of the MCL, and by Peters *et al.* (2010) who propose Bayesian and classical bootstrap methods applied to all the Chain Ladder parameters. In our paper, the distribution-free aspect is treated by using non-parametric kernel methods. Another difference between our approach and the others is that we do not impose the same stochastic behaviour across the accident years.

2.2. Dependence

The original PIC method has been extended in order to take into account the dependence between the observations, in particular, the dependence between the rates of increase of the payments and of the incurred losses in a given accident year. Happ and Wüthrich (2013) assumed that the incurred losses increments are correlated with the claims payments at present and future dates. Peters *et al.* (2014) adopted a copula approach. In our approach, the variance–covariance matrix of all the rates of increase of payments and incurred losses of a given accident year is full, in the sense that all the correlations (between the cumulated payments and the incurred losses at a same or a different development year) are allowed to be non-zero and the structure of the variance–covariance matrix is a natural consequence of the equality between the payments and incurred losses at the last development year (see Proposition 3.1).

2.3. Ultimate development year

The ultimate development year is unknown. This problem has been treated in the Chain Ladder context by the introduction of "tail development factors" (see Boor, 2006). Merz and Wüthrich (2010) also considered the tail development factors within their PIC method. In our paper, we adopt a different approach. The ultimate development date is considered as an additional parameter and an estimation method is proposed.

2.4. Reserve risk measures

Several papers used the mean square error prediction (MSEP) as a risk measure. In Happ and Wüthrich (2013), the MSEP is applied to the total ultimate claims, whereas in Dahms *et al.* (2009) and Happ *et al.* (2012), it is applied to the CDR. In Peters *et al.* (2010), both the MSEP and the VaR of the total ultimate claims are considered. In our paper, we follow the recommendation of Solvency II and we consider the VaR of the CDR's. Moreover, this can be done in a distribution-free framework.

2.5. Sensitivity of the CDR VaR

In our paper, we study two points which are not considered in the literature: the sensitivity of the CDR VaR's to non-Gaussianity and to parameter estimation updating. The application on real data show that this sensitivity is huge and, therefore, should not be neglected.

3. A GAUSSIAN CIP MODEL

The main objective of this paper is to propose a semi-parametric model that makes no assumptions about the probability distributions but only on their first- and second-order moments. However, since the Gaussian assumption is very often retained in the literature, we start with a Gaussian version of our model, which will be a useful benchmark.

3.1. Notations

We denote, respectively, by $P_{i,j}$ and $I_{i,j}$, the cumulated payments and incurred losses for accident year i and development year j. The calendar year is $i+j$. We also use the following notations:

$$
\begin{aligned}
X_{1,i,j} &= \log\ P_{i,j},\\
X_{2,i,j} &= \log\ I_{i,j},\\
X_{i,j} &= \begin{pmatrix} X_{1,i,j} \\ X_{2,i,j} \end{pmatrix},\\
Y_{1,i,j} &= X_{1,i,j} - X_{1,i,j-1} = \log\ \frac{P_{i,j}}{P_{i,j-1}},\\
Y_{2,i,j} &= X_{2,i,j} - X_{2,i,j-1} = \log\ \frac{I_{i,j}}{I_{i,j-1}},\\
Y_{i,j} &= \begin{pmatrix} Y_{1,i,j} \\ Y_{2,i,j} \end{pmatrix}.
\end{aligned}
$$

We assume that $X_{i,j}$ is observed for

$$
\begin{aligned}
& i = 1, \ldots, n,\\
& j = 0, \ldots, n-1,\\
& 1 \le i + j \le n,
\end{aligned}
$$

and, consequently, $Y_{i,j}$ is observed for

$$
\begin{aligned}
& i = 1, \ldots, n-1,\\
& j = 1, \ldots, n-1,\\
& 2 \le i + j \le n.
\end{aligned}
$$

A key assumption, throughout the paper, is that there is an ultimate development year $N \ge n-1$, in general not observed, such that $X_{1,i,N} = X_{2,i,N}$ for all i,

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

and the models proposed will have to satisfy this constraint. We also introduce the notations:

$$\begin{array}{ll} Y_i = (Y'_{i,1}, \ldots, Y'_{i,n-i})' & \text{of size } 2(n-i), \\ \widetilde{Y}_i = (Y'_{i,1}, \ldots, Y'_{i,N})' & \text{of size } \ 2N, \\ \widetilde{Y}_{1,i} = (Y_{1,i,1}, \ldots, Y_{1,i,N})' & \text{of size } \ N, \\ \widetilde{Y}_{2,i} = (Y_{2,i,1}, \ldots, Y_{2,i,N})' & \text{of size } \ N. \end{array}$$

3.2. A conditional Gaussian model

A first CIP model is obtained by starting from a Gaussian model and then imposing the conditioning constraints:

$$X_{1,i,N} = X_{2,i,N}, i = 1, \ldots, n-1.$$

More precisely, we assume that $X_{i,0}$ is fixed and, in a first step, we introduce the model:

$$Y_{i,j} = m(i, j, \theta) + \xi_{i,j}, \tag{1}$$

where the $m(i, j, \theta)$ are bidimensional deterministic functions and the $\xi_{i,j}$ are bidimensional vectors following independently the Gaussian distribution $N[0, \Omega(i, j, \theta)]$, where θ is an unknown vector of parameters. Note that $\Omega(i, j, \theta)$ is not assumed to be diagonal and, therefore, at this stage, a correlation between the two components of $\xi_{i,j}$ is allowed.

Then the probabilistic structure of the model is modified by introducing the constraints:

$$X_{1,i,N} = X_{2,i,N}, i = 1, \ldots, n-1. \tag{2}$$

In other words, we assume that the $\widetilde{Y}_i$, $i = 1, \ldots, n-1$ are independently distributed and that the distribution of $\widetilde{Y}_i$ is the conditional distribution obtained from the initial Gaussian model (1) by imposing $X_{1,i,N} = X_{2,i,N}$ or, equivalently, $d' X_{i,N} = 0$ with $d' = (1, -1)$. Finally, we obtain the following distribution of the vector $\widetilde{Y}_i$.

Propositon 3.1. *The conditional distribution of $\widetilde{Y}_i$ given $d' X_{i,N} = 0$ is the Gaussian distribution:*

$$N\left(\widetilde{m}_i - \frac{\widetilde{c}_i a_i}{b_i}, \widetilde{\Omega}_i - \frac{\widetilde{c}_i \widetilde{c}'_i}{b_i}\right),$$

where

$$\widetilde{m}_i = [m'(i, 1, \theta), \ldots, m'(i, N, \theta)]',$$

$$\widetilde{c}_i = [d'\Omega(i, 1, \theta), \ldots, d'\Omega(i, N, \theta)]',$$

$$\widetilde{\Omega}_i = \begin{bmatrix} \Omega(i, 1, \theta) & & 0 \\ & \cdots & \\ 0 & & \Omega(i, N, \theta) \end{bmatrix},$$

$$a_i = d' X_{i,0} + d' \sum_{j=1}^{N} m(i, j, \theta),$$

$$b_i = d' \sum_{j=1}^{N} \Omega(i, j, \theta) d$$

[*for sake of notational simplicity, we use the notations* $\tilde{m}_i$ *instead of* $\tilde{m}_i(\theta)$, $\tilde{c}_i$ *instead of* $\tilde{c}_i(\theta)$, $\tilde{\Omega}_i$ *instead of* $\tilde{\Omega}_i(\theta)$, a_i *instead of* $a_i(\theta)$ *and* b_i *instead of* $b_i(\theta)$].

Proof. see Appendix A. ■

Note that using the notation $F_N = (I_2, \ldots, I_2)'$ where the identity matrix of size 2 is repeated N times, we have

$$\widetilde{c}_i = \widetilde{\Omega}_i F_N d,$$

$$a_i = d'(X_{i,0} + F_N' \widetilde{m}_i),$$

$$b_i = d' F_N' \widetilde{\Omega}_i F_N d = d' F_N' \widetilde{c}_i.$$

It is important to stress that the variance–covariance matrix of $\tilde{Y}_i$, namely $\tilde{\Sigma}_i = \tilde{\Omega}_i - \frac{\tilde{c}_i \tilde{c}_i'}{b_i}$ is full, in other words, all the components of $\tilde{Y}_i$ are correlated. This correlation structure is introduced in a natural way by constraints (2) imposing equality of the cumulated payments and the incurred losses at the ultimate development year.

We easily deduce the conditional distribution of the observed vector Y_i.

Corollary 3.1. *The conditional distribution of* Y_i *given* $d' X_{i,N} = 0$ *is the Gaussian distribution :* $N\left(m_i - \frac{c_i a_i}{b_i}, \Omega_i - \frac{c_i c_i'}{b_i}\right)$, where $m_i = [m'(i, 1, \theta), \ldots, m'(i, n - i, \theta)]'$ $c_i = [d'\Omega(i, 1, \theta), \ldots, d'\Omega(i, n - i, \theta)]'$

$$\Omega_i = \begin{bmatrix} \Omega(i, 1, \theta) & & 0 \\ & \cdots & \\ 0 & & \Omega(i, n - i, \theta) \end{bmatrix}.$$

Proof. We just have to take the marginal distribution of the first $n - i$ components of the joint distribution given in Proposition 3.1. ■

Note that the Gaussian distribution of Proposition 3.1 is degenerated since the components of $\widetilde{Y}_i$ have to satisfy the linear constraint:

$$d' X_{i,0} + d' \sum_{j=1}^{N} Y_{i,j} = 0$$

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

or

$$d'(X_{i,o} + F'_N \widetilde{Y}_i) = 0.$$

The matrix

$$\widetilde{\Omega}_i - \frac{\widetilde{c}_i \widetilde{c}'_i}{b_i}$$

is of rank $2N - 1$. However, as soon as n is strictly smaller than $N + 1$ the variance–covariance matrix of Y_i namely

$$\Omega_i - \frac{c_i c'_i}{b_i}$$

is of full rank $2(n - i)$ for all i including $i = 1$.

3.3. Estimation of a Gaussian CIP model

As soon as the functions $m(i, j, \theta)$ and $\Omega(i, j)$ have been specified (see Section 5 for a discussion of these specifications), the parameter θ can be estimated by the ML method. Indeed from Corollary 3.1, we deduce that the log-likelihood function of the model is

Propositon 3.2.

$$L_n(\theta) = -\frac{1}{2} \sum_{i=1}^{n-1} \left[\log det \Sigma_i(\theta) + (y_i - \mu_i(\theta))' \Sigma_i^{-1}(\theta)(y_i - \mu_i(\theta))\right]$$

with

$$\mu_i(\theta) = m_i(\theta) - \frac{c_i(\theta) a_i(\theta)}{b_i(\theta)},$$

$$\Sigma_i(\theta) = \Omega_i(\theta) - \frac{c_i(\theta) c'_i(\theta)}{b_i(\theta)}.$$

Proof. It is a direct consequence of the expression of the probability density function of a multivariate Gaussian distribution. ■

Moreover, the computation of $\Sigma_i^{-1}(\theta)$ is simple thanks to the following proposition (omitting θ for notational simplicity).

Propositon 3.3.

$$\Sigma_i^{-1} = \Omega_i^{-1} + \frac{\Omega_i^{-1} c_i c'_i \Omega_i^{-1}}{b_i - c'_i \Omega_i^{-1} c_i}$$

with

$$\Omega_i^{-1} = \begin{pmatrix} \Omega_{i,1}^{-1} & & 0 \\ & \cdots & \\ 0 & & \Omega_{i,n-i}^{-1} \end{pmatrix}$$

[*where* $\Omega_{i,j}^{-1}$ *is a notation for* $\Omega^{-1}(i, j, \theta)$].

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

Proof. See Appendix B. ■

In particular, the previous proposition implies that the term $(y_i - \mu_i)'\Sigma_i^{-1}(y_i - \mu_i)$ in the log-likelihood is simply

$$\sum_{j=1}^{n-i}\left[(y_{i,j}-\mu_{i,j})'\Omega_{i,j}^{-1}(y_{i,j}-\mu_{i,j}) + \frac{[(y_{i,j}-\mu_{i,j})'\Omega_{i,j}^{-1}c_{i,j}]^2}{b_i - \sum_{j=1}^{n-i} c'_{i,j}\Omega_{i,j}^{-1}c_{i,j}}\right]. \quad (3)$$

Starting values for the parameters appearing in the $m(i, j, \theta)$ can be obtained by the ordinary least squares (OLS) method and from the residuals of this method for the parameters appearing in the $\Omega(i, j, \theta)$ (see Section 5).

The ML estimator of θ will be denoted by $\widehat{\theta}_n$. Note that it is based on $(n-1)+(n-2)+\cdots+1 = \frac{n(n-1)}{2}$ observations $Y_{i,j}$ of size 2.

The whole testing and confidence region methods based on ML estimators apply. In particular, the variance–covariance matrix of $\widehat{\theta}_n$ can be approximated by (see e.g. Gourieroux and Monfort (1996) chapter 7)

$$-\left[\frac{\partial^2 L_n(\widehat{\theta}_n)}{\partial\theta\partial\theta'}\right]^{-1}. \quad (4)$$

In order to test a constraint $g(\theta) = 0$ where g is a multivariate function of size r, we can use the Wald statistic:

$$\xi_n^w = -g'(\hat{\theta}_n)\left[\frac{\partial g}{\partial\theta'}(\hat{\theta}_n)\left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial g'(\hat{\theta}_n)}{\partial\theta}\right]^{-1} g(\hat{\theta}_n)$$

and reject the constraint if ξ_n^w is larger than $\chi^2_{1-\alpha}(r)$, the quantile of order $1-\alpha$ of the χ^2 distribution with r degrees of freedom, α being the level of the test. In particular, if we want to test that the kth component θ_k of θ is equal to zero, we can use the critical region:

$$\frac{\hat{\theta}^2_{k,n}}{\hat{\sigma}^2_k} \geq \chi^2_{1-\alpha}(1),$$

where $\hat{\sigma}^2_k$ is the kth diagonal term of $-\left[\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right]^{-1}$, or equivalently

$$\frac{|\hat{\theta}_{k,n}|}{\hat{\sigma}_k} \geq u_{1-\alpha/2},$$

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

where $u_{1-\alpha|2}$ is the $1-\alpha/2$ quantile of $N(0, 1)$, the standard Gaussian distribution. $\frac{|\hat{\theta}_{kn}|}{\hat{\sigma}_k}$ is called the t-ratio. If we take $\alpha = 5\%$, $u_{1-\alpha/2} = 1.96$, and therefore we reject $\theta_k = 0$ if the t ratio is larger than 1.96.

3.4. Tail development

As mentioned in Section 3.1, the ultimate development year N is, in general, larger than the latest development year j where observations of Y_i are available, namely $j = n-1$. It is the so called "tail development" problem. In the context of Chain Ladder approaches, this problem has been reduced to the computation of an ultimate development factor called "tail development factor." A review of these approaches is available in Boor (2006). The tail development problem has also been considered by Merz and Wuthrich (2010) within their PIC method; in particular, their Bayesian approach allows for a tail development factor covering several development periods beyond the last column of the claim development triangle.

In our approach, we consider N as an unknown parameter. It is clear that the previous log-likelihood depends on N through the $a_i^{\prime}s$ and the $b_i^{\prime}s$. Therefore, we can use recent results on the estimation of discrete parameter models (see Choirat and Seri 2012) showing that maximizing the log-likelihood function with respect to all the parameters, including N, provides a consistent estimator of N. Indeed the results in Choirat, Seri paper are valid for the m-estimator family, which contains the ML estimators. This gives us not only coherent estimates of the ultimate values $P_{i,N} = I_{i,N}$, for any i, but also a consistent estimate of the length of the tail development period.

3.5. Prediction

Once the parameters are estimated, we have to predict $Y_i^* = (Y_{i,n-i+1}^{\prime}, \ldots, Y_{i,N}^{\prime})^{\prime}$, or, equivalently, $X_i^* = (X_{i,n-i+1}^{\prime}, \ldots, X_{i,N}^{\prime})^{\prime}$ for each i, given the observations $X_{i,0}, Y_{i,1}, \ldots, Y_{i,n-i}$ or, equivalently $X_i = (X_{i,0}^{\prime}, \ldots, X_{i,n-i}^{\prime})^{\prime}$.

The conditional distribution of X_i^* given X_i (without conditioning by $d^{\prime}X_{i,N} = 0$) is the same as the conditional distribution of X_i^* given $X_{i,n-i}$ since for any $k \in \{1, \ldots, N-n+i\}$:

$$X_{i,n-i+k} = X_{i,n-i} + \sum_{j=1}^{k} Y_{i,n-i+j}$$

and the $Y_{i,n-i+k}$ are independent of X_i. Therefore, the conditional distribution of X_i^* given X_i and $d^{\prime}X_{i,N} = 0$ is the same as the conditional distribution of X_i^* given $X_{i,n-i}$ and $d^{\prime}X_{i,N} = 0$ (since $d^{\prime}X_{i,N}$ is function of X_i^*). This implies that we have to solve the same problem as in Section 3.2, just replacing $X_{i,0}$ by $X_{i,n-i}$ and $\widetilde{Y}_i$ by Y_i^*, and we get the following proposition.

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

Propositon 3.4. The conditional distribution of Y_i^* given $X_{i,n-i}$ and $d' X_{i,N} = 0$ is the Gaussian distribution:

$$N\left(m_i^* - \frac{c_i^* a_i^*}{b_i^*}, \Omega_i^* - \frac{c_i^* c_i'^*}{b_i^*}\right)$$

with

$$m_i^* = [m'(i, n-i+1, \theta), \ldots, m'(i, N, \theta)]',$$
$$c_i^* = [d'\Omega(i, n-i+1, \theta), \ldots, d'(\Omega_i, N, \theta)]',$$
$$\Omega_i^* = \begin{bmatrix} \Omega(i, n-i+1, \theta) & & 0 \\ & \ldots & \\ 0 & & \Omega(i, N, \theta) \end{bmatrix},$$
$$a_i^* = d'\left(X_{i,n-i} + \sum_{j=n-i+1}^{N} m(i, j, \theta)\right),$$
$$b_i^* = d' \sum_{j=n-i+1}^{N} \Omega(i, j, \theta) d.$$

We also will use the notations

$$\mu_i^* = m_i^* - \frac{c_i^* a_i^*}{b_i^*}, \Sigma_i^* = \Omega_i^* - \frac{c_i^* c_i'^*}{b_i^*}$$

and μ_i^*, of size $2(N-n+i)$, is partitioned into $N-n+i$ bidimensional vectors, in the following way:

$$\mu_i^* = [\mu^{*'}(i, n-i+1, \theta), \ldots, \mu^{*'}(i, N, \theta)]'.$$

The best prediction of Y_i^* (minimizing the mean square prediction error) is μ_i^* and the best prediction of X_{n-i+k}, $k = \{1, \ldots, N-n+i\}$ is $X_{i,n-i} + \sum_{j=1}^{k} \mu^*(i, n-i+j, \theta)$.

3.6. Claim development results (CDR)

Denoting by E_n the conditional expectation operator with respect to the true conditional distribution given the information at the calendar date n :

$$\mathcal{J}_n = \{X_{i,0}, i = 1, \ldots, n-1, Y_{i,j}, i = 1, \ldots, n-1, j = 1, \ldots, n-i\}$$

in which the true value of the parameter θ is evaluated at $\widehat{\theta}_n$, the CDR for the accounting calendar period $(n, n+1)$ and accident year i is

$$\text{CDR}_i(n+1) = E_n(X_{1,i,N}) - E_{n+1}(X_{1,i,N}) \tag{5}$$

or, equivalently

$$\mathrm{CDR}_i(n+1) = E_n(X_{2,i,N}) - E_{n+1}(X_{2,i,N}), \tag{6}$$

since in our model, we automatically have $X_{1,i,N} = X_{2,i,N}$.

It is important to note that in the E_{n+1} operator the value of θ is set at $\hat{\theta}_{n+1}$, in other words the updating of the estimation of θ is taken into account.

Choosing $X_{1,i,N}$, we can write

$$X_{1,i,N} = X_{1,i,n-i} exp\left(\sum_{j=n+1-i}^{N} Y_{1,i,j}\right) = X_{1,i,n-i}\exp(f_i' Y_i^*), \tag{7}$$

where f_i' is the row vector of size 2 $(N - n + i)$ equal to $(1, 0, 1, 0, \ldots, 1, 0)$ picking the components $Y_{1,i,j}$, $j = n - i + 1, \ldots, N$ in Y_i^*.

Therefore, the true conditional expectation of $X_{1,i,N}$ is

$$X_{1,i,n-i}\, exp[f_i' \mu_i^*(\theta_0) + 1/2 f_i' \Sigma_i^*(\theta_0) f_i], \tag{8}$$

where θ_0 is the true value of θ.

Replacing θ by the ML estimator $\widehat{\theta}_n$, we get

$$E_n(X_{1,i,N}) = X_{1,i,n-i}\exp\left[f_i' \mu_i^*(\widehat{\theta}_n) + 1/2 f_i' \Sigma_i^*(\widehat{\theta}_n) f_i\right]. \tag{9}$$

Similarly we have

$$E_{n+1}(X_{1,i,N}) = X_{1,i,n-i+1}\exp\left[f_i'^* \mu_i^{**}(\widehat{\theta}_{n+1}) + 1/2 f_i'^* \Sigma_i^{**}(\widehat{\theta}_{n+1}) f_i^*\right], \tag{10}$$

where f_i^* and μ_i^{**} are obtained from f_i and μ_i^*, respectively, by deleting the first two components and Σ_i^{**} is obtained from Σ_i^* by deleting the first two rows and the first two columns.

$X_{1,i,n-i+1}$ is random at date n and is equal to $X_{1,i,n-i}\exp(Y_{1,i,n-i+1})$ with

$$Y_{1,i,n-i+1} = \mu_{1,i}^*(\theta_0) + \sigma_{1,i}^*(\theta_0)\varepsilon_{1,i,n-i+1},$$

where $\mu_{1,i}^*(\theta_0)$ is the first component of $\mu_i^*(\theta_0)$, $\sigma_{1,i}^*(\theta_0)$ the square root of the $(1, 1)$ term of $\Sigma_i^*(\theta_0)$ and $\varepsilon_{1,i,n-i+1}$ is following $N(0, 1)$.

It is natural to view $\mathrm{CDR}_i(n+1)$ from the calendar date n and, therefore, to replace θ_0 by $\widehat{\theta}_n$ in the expression above of $Y_{1,i,n-i+1}$.

Finally, we get the estimation of $\mathrm{CDR}_i(n+1)$:

$$\widehat{\mathrm{CDR}}_i(n+1) = X_{1,i,n-i}\left\{exp\left[f_i' \mu_i^*(\widehat{\theta}_n) + \frac{1}{2} f_i' \Sigma_i^*(\widehat{\theta}_n) f_i\right]\right.$$

$$\left. -\exp[\mu_{1,i}^*(\widehat{\theta}_n) + \sigma_{1,i}^*(\widehat{\theta}_n)\varepsilon_{1,i,n-i+1} + f_i'^* \mu_i^{**}(\widehat{\theta}_{n+1}) + 1/2 f_i'^* \Sigma_i^{**}(\widehat{\theta}_{n+1}) f_i^*]\right\}. \tag{11}$$

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

At date n, there are two sources of randomness in $\widehat{\text{CDR}}_i(n+1)$. Indeed $\widehat{\text{CDR}}_i(n+1)$ is random through $\varepsilon_{1,i,n-i+1}$ and through the $\varepsilon_{1,k,n-k+1}$, $k = 1, \ldots, n-1$ (containing $\varepsilon_{1,i,n-i+1}$) and $\varepsilon_{2,k,n-k+1}$ appearing in the new observations $Y_{i,n-i+1}$ at calendar date $n+1$ which are used in the updated estimation $\hat{\theta}_{n+1}$ of θ. The global CDR is estimated by $\widehat{\text{CDR}}(n+1) = \sum_{i=1}^{n} \widehat{\text{CDR}}_i(n+1)$. The $\widehat{\text{CDR}}_i(n+1)'s$ and $\widehat{\text{CDR}}(n+1)$ are random variables which can be simulated, as explained in the next section.

3.7. Value at Risk (VaR) of the $\widehat{\text{CDR}}_i(n+1)$

The VaR $\text{VaR}_i(\alpha)$ associated with $\widehat{\text{CDR}}_i(n+1)$, or rather with the under-provisioning measure $-\widehat{CDR}_i(n+1)$, is defined by

$$P[-\widehat{\text{CDR}}_i(n+1) < \text{VaR}_i(\alpha)] = \alpha,$$

where α is close to 1, for instance 0.995.

If we do not take into account the updating of $\widehat{\theta}_n$ and set $\widehat{\theta}_{n+1} = \widehat{\theta}_n$, the only random term in (11) is $\varepsilon_{1,i,n-i+1}$ distributed as $N(0, 1)$. In other words, with obvious notations, $\widehat{\text{CDR}}_i(n+1)$ is of the form $\beta_i - \gamma_i \exp(\delta_i \varepsilon_{1,i,n-i+1})$ with $\beta_i > 0, \gamma_i > 0, \delta_i > 0$.

The $\text{VaR}_i(\alpha)$ is easily seen to be

$$\gamma_i \exp[\delta_i \Phi(\alpha)] - \beta_i. \tag{12}$$

If we want to take into account the updating of $\widehat{\theta}_n$ into $\widehat{\theta}_{n+1}$, we might use the Newton–Raphson approximation:

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n - \left[\frac{\partial^2 L_n(\widehat{\theta}_n)}{\partial\theta\partial\theta'}\right]^{-1} \frac{\partial L_{n+1}}{\partial\theta}(\widehat{\theta}_n), \tag{13}$$

where $\frac{\partial^2 L_n}{\partial\theta\partial\theta'}(\widehat{\theta}_n)$ is a by-product of the estimation procedure and $\frac{\partial L_{n+1}}{\partial\theta}(\widehat{\theta}_n)$ can be computed numerically as a function of the $\varepsilon_{k,n-k+1}$, $k = 1, \ldots, n-1$.

Then $\text{VaR}_i(\alpha)$ can be evaluated by simulation. More precisely, let us consider M simulations of $\widehat{\text{CDR}}_i(n+1)$ and let us order them in increasing order, then $-\text{VaR}_i(\alpha)$ is taken equal to the value with index $[M\alpha]$ (where [.] is a notation for the integer).

Note that for the computation of the $\text{VaR}(\alpha)$ of the global $\widehat{\text{CDR}}(n+1) = \sum_{i=1}^{n} \widehat{\text{CDR}}_i(n+1)$, such a simulation method is required even when we do not update $\widehat{\theta}_n$.

3.8. Comparison with other models based on the normal or log-normal distributions

Although the main objective of this paper is to propose the semi-parametric approach developed in Section 4, it is of some interest to compare the Gaussian (or normal) version of our method with other methods using either the normal or the log-normal distribution. It turns out that important difference with most methods is the use of conditional distributions imposing equality of the cumulated payments and incurred losses at an ultimate development year. This conditioning approach has been used only by Posthuma *et al.* (2008). Another similarity between this paper and our work is the fact that, in both models, the number of parameters does not depend on the number of observations, thus avoiding the incidental problem and allowing for the use of the asymptotic results of the ML theory in particular the use of Wald and t-ratio tests. Finally, in both models, we adopt a classical approach and not a Bayesian approach. There are, however, several important differences between both approaches. In our approach, we work with the logarithm of the variables and, therefore the variables themselves have a log-normal distribution, not a normal distribution. We introduce a correlation between the variables before the conditioning operation, and the correlation structure is then enriched by the conditioning operation leading to a situation where all the variables of a given accident year are correlated (see the variance–covariance matrix in proposition 3.1). The ultimate date at which the cumulated payments and the incurred losses are equal is unknown and estimated. We also consider the computation of the CDRs of their VaRs.

4. A SEMI-PARAMETRIC CLASS OF CIP MODELS

4.1. Semi-parametric models

In the Gaussian model, the vectors Y_i, of size $2(n-i)$, follow independently the distribution (see Corollary 3.1):

$$N(\mu_i, \Sigma_i)$$

with

$$\mu_i = m_i - \frac{c_i a_i}{b_i}, \Sigma_i = \Omega_i - \frac{c_i c_i'}{b_i}.$$

Let us denote by T_i the lower triangular matrix or Cholesky matrix, such that $\Sigma_i = T_i T_i'$ (imposing positive diagonal terms for T_i implies its uniqueness), we can write

$$Y_i = \mu_i + T_i \varepsilon_i, \tag{14}$$

$$\varepsilon_i \sim N(0, I_{2(n-i)}) \quad \text{or}$$

$$\varepsilon_{i,k} \sim IIN(0, 1),\ k = 1, \ldots, 2(n-i).$$

A natural extension of this model consists in still assuming $Y_i = \mu_i + T_i \varepsilon_i$ but only imposing that the $\varepsilon_{i,k}$ are identically, independently distributed with

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

zero mean and unit variance:

$$Y_i = \mu_i + T_i \varepsilon_i, \qquad (15)$$
$$\varepsilon_{i,k} \sim II(0, 1).$$

In other words, we no longer assume that the $\varepsilon_{i,k}$ are Gaussian and we do not make any assumption about their common distribution. Note however that the first- and second-order moments of Y_i are still μ_i and Σ_i. In particular, all the components of Y_i are correlated. The model becomes semi-parametric and the ML method is no longer available.

4.2. Pseudo maximum likelihood estimation of θ

Gourieroux *et al.* (1984) have shown that if only the conditional mean of a random variable is parametrically specified and if the parameters are estimated by maximizing a pseudo likelihood based on a family of instrumental density functions indexed by their means, this estimation is consistent and asymptotically Gaussian if and only if the instrumental family is exponential affine. Similarly, if only the conditional means and the conditional variance–covariance matrices are parametrically specified, a PML estimator is consistent and asymptotically normal if, and only if, the instrumental family is exponential quadratic. The Gaussian family is such an exponential quadratic family of density functions. This implies that the PML estimator of θ obtained by maximizing $L_n(\theta)$ given in Proposition 3.2 is still consistent and asymptotically normal for any distribution of the $\varepsilon'_{i,j}s$. The only modification is the estimation of the variance–covariance matrix of the estimator of θ that becomes

$$\left[\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right]^{-1} \sum_{i=1}^{n-1} \frac{\partial \log f_i(\hat{\theta}_n)}{\partial\theta} \frac{\partial \log f_i(\hat{\theta}_n)}{\partial\theta'} \left[\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right]^{-1}, \qquad (16)$$

where $-2 \ \log \ f_i(\theta)$ is given by formula (3). Formula (16) is sometimes called the "sandwich" formula. The testing procedures described in Section 3.3 remain valid if the estimation of the variance–covariance matrix of the estimator of θ, previously given by formula (4), is replaced by formula (16). Finally, note that the PML method is a particular m-estimation method and, therefore, the results by Choirat and Seri (2012) valid for the m-estimators, can be used in the PML context, implying the consistency of the PML estimator of N.

4.3. Non-parametric estimation of the distribution of $\varepsilon_{i,k}$.

We can then estimate the ε_i by

$$\widehat{\varepsilon}_i = \widehat{T}_i^{-1}(Y_i - \widehat{\mu}_i), \qquad (17)$$

where $\widehat{T}_i$ and $\widehat{\mu}_i$ are T_i and μ_i evaluated at $\widehat{\theta}_n$.

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

The unknown distribution of the $\varepsilon'_{i,k}s$ can be estimated by the Gaussian kernel method and we get the following mixture of Gaussian distributions:

$$\frac{1}{n(n-1)h}\sum_{i=1}^{n-1}\sum_{k=1}^{2(n-i)}\varphi\left(\frac{\varepsilon-\widehat{\varepsilon}_{i,k}}{h}\right),$$

where φ is the p.d.f of $N(0, 1)$ and h is equal to $[n(n-1)]^{-1/5}$ according to Silverman's rule. The mean and the variance of this mixture of Gaussian distributions are

$$\overline{\varepsilon} = \frac{1}{n(n-1)}\sum_{i,k}\widehat{\varepsilon}_{i,k},$$

$$\overline{\sigma}^2 = h^2 + \frac{1}{n(n-1)}\sum_{i,k}\widehat{\varepsilon}^2_{i,k} - \overline{\varepsilon}^2,$$

and, in order to get a distribution zero mean and unit variance, we can use

$$\frac{\overline{\sigma}}{n(n-1)h}\sum_{i,k}\varphi\left(\frac{\overline{\sigma}\varepsilon+\overline{\varepsilon}-\widehat{\varepsilon}_{i,k}}{h}\right).$$

Also, note that a preliminary test of Gaussianity of the $\varepsilon'_{i,k}s$ can be made with the Jarque–Bera procedure which rejects the Gaussianity at level α (for instance, $\alpha = 5\%$) if

$$n(n-1)\left(\frac{S^2}{6}+\frac{(K-3)^2}{24}\right) \geq \chi^2_{1-\alpha}(2), \tag{18}$$

where S and K are, respectively, the empirical skewness and kurtosis of the $\widehat{\varepsilon}_{i,k}$.

4.4. The CDR$_i$ and their VaR

From Proposition 3.4, we know that, in the Gaussian case, the conditional distribution of $Y_i^* = (Y'_{i,n+1-i}, \ldots, Y'_{i,N})'$ given $X_{i,n-i}$ and $d'X_{i,N} = 0$ is $N(\mu_i^*, \Sigma_i^*)$ with

$$\mu_i^* = m_i^* - \frac{c_i'^* a_i^*}{b_i^*},\ \Sigma_i^* = \Omega_i^* - \frac{c_i^* c_i'^*}{b_i^*}.$$

Denoting by T_i^* the Cholesky matrix satisfying $T_i^* T_i'^* = \Sigma_i^*$, we have

$$Y_i^* = \mu_i^* + T_i^* \varepsilon_i^*, \tag{19}$$

where the components $\varepsilon_{i,k}^*$ of ε_i^* follow independently $N(0, 1)$. In the general case, we can make the assumption that these components $\varepsilon_{i,k}^*$ follow

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

independently a distribution estimated by the one obtained in Section 4.3. At this stage, it is important to stress the following property.

Propositon 4.1. *For any distribution of the* $\varepsilon^*_{i,k}$*, the model* $Y^*_i = \mu^*_i + T^*_i \varepsilon^*_i$ *implies* $X_{1,i,N} = X_{2,i,N}$.

Proof. Since the model $Y^*_i = \mu^*_i + T^*_i \varepsilon^*_i$ implies, for any distribution of the $\varepsilon^*_{i,k}$, the same first- and second-order moments of Y^*_i and X^*_i as in the Gaussian model, we have in particular $E(X_{1,i,N} - X_{2,i,N}) = 0$ and $V(X_{1,i,N} - X_{2,i,N}) = 0$ and therefore $X_{1,i,N} = X_{2,i,N}$ for any distribution of the $\varepsilon^*_{i,k}$. ■

Since the conditional expectation of Y^*_i given $X_{i,n-i}$ and $d' X_{i,N} = 0$, remains equal to μ^*_i, the best prediction of $X_{i,n-i+k}, k \in \{1, \ldots N - n + i\}$, remains $X_{i,n-i} + \sum_{j=1}^{k} \mu^*(i, n - i + j; \theta)$.

The $\text{CDR}_i(n+1)$ is

$$\text{CDR}_i(n+1) = X_{1,i,n-i} E_n[\exp(f'_i Y^*_i)] - X_{1,i,n-i+1} E_{n+1}[exp(f'^*_i Y^{**}_i)]$$

with

$$Y^{**}_i = (Y'_{i,n-i+2}, \ldots, Y'_{i,N})'$$

or

$$\text{CDR}_i(n+1) = X_{1,i,n-i}[E_n \exp(f'_i Y^*_i) - \exp(Y_{1,i,n-i+1}) E_{n+1} \exp(f'^*_i Y^{**}_i)]. \quad (20)$$

From (19), we get

$$Y_{1,i,n-i+1} = \mu^*_{i,1} + T^*_{i,11}\varepsilon^*_{i,1}$$

and replacing the true value of θ_0 appearing in $\mu^*_{i,1}$ and $T^*_{i,11}$by $\widehat{\theta}_n$, we get

$$\hat{Y}_{1,i,n-i+1} = \widehat{\mu}^*_{i,1} + \widehat{T}^*_{i,11}\varepsilon^*_{i,1}$$

and

$$\widehat{\text{CDR}}_i(n+1) = X_{i,n-i}[E_n \exp(f'_i Y^*_i) - \exp(\widehat{\mu}^*_{i,1} + \hat{T}^*_{i,11}\varepsilon^*_{i,1}) E_{n+1} \exp(f'^*_i Y^{**}_i)].$$

If we do not take into account the estimation updating, we can easily simulate $\widehat{\text{CDR}}_i(n+1)$ by simulating $\varepsilon^*_{i,1}$ in the distribution estimated in Section 4.3 and by computing both expectations by Monte Carlo using the values $\widehat{\theta}_n$ in the relevant components of μ^*_i and T^*_i.

If we want to take into account the estimation updating, for each simulation of $Y_{i,n+1-i}$ based on

$$Y_{i,n-i+1} = \begin{pmatrix} \mu^*_{i,1} \\ \mu^*_{i,2} \end{pmatrix} + \begin{pmatrix} T^*_{i,11} & 0 \\ T^*_{i,21} & T^*_{i,22} \end{pmatrix} \begin{pmatrix} \varepsilon^*_{i,1} \\ \varepsilon^*_{i,2} \end{pmatrix},$$

we must update $\widehat{\theta}_n$ into $\widehat{\theta}_{n+1}$ and, then compute the second expectation in (20) by Monte Carlo, replacing θ_0 by $\widehat{\theta}_{n+1}$ in the equations

$$Y_i^{**} = \mu_i^{**} + T_i^{**} \varepsilon_i^{**}.$$

The estimations of the $\text{VaR}_i(\alpha)'s$ and of the global $\text{VaR}(\alpha)$ are obtained from the empirical quantiles of M simulations of the $\widehat{\text{CDR}}_i(n+1)'s$ and of

$$\widehat{\text{CDR}}(n+1) = \sum_{i=1}^{n} \widehat{\text{CDR}}_i(n+1).$$

4.5. Comparison with other distribution-free methods

The distribution-free methods appearing in the literature are bootstrap methods. Liu and Verrall (2010) adopt a classical bootstrap approach based on the MCL taking into account both the paid and incurred claim triangles. Peter *et al.* (2010) propose a Bayesian bootstrap approach based on the ABC (Approximate Bayesian Computation) method using MCMC (Monte Carlo Markov Chain) techniques and they consider only one kind of claims (payments or claims incurred). By contrast, our distribution-free method considers both kinds of claims and uses non-parametric kernel techniques, based on first step estimations of the parameters appearing in the first- and second-order moments estimated by the PML method. Moreover, none of the bootstrap methods mentioned above consider the computation of the CDR's and of their VaR's and the estimation of the ultimate date.

5. An Application

We consider cumulated payments and incurred claims corresponding to a line of business Motor Body Liability-Insurance of 14 accident years (the unit is 10^3 euros) [see Appendix C]. This line of business is highly volatile and therefore, not easy to model.

5.1. Estimation of the parameters

We begin with separate modelling of the rates of growth of cumulated payments and of incurred claims. For each variable, we estimate by the non-linear least square method a mean function, i.e. the corresponding component of $m(i, j, \theta)$, and a variance function i.e. the corresponding diagonal term of $\Omega(i, j, \theta)$. The mean function is assumed to be an affine function, with unknown coefficients (the components of θ) of basic functions of i and j, namely the identity function, the square function, the logarithmic function and the exponential. These mean functions are also assumed to be equal to zero if j is larger than a threshold J.

TABLE 1

CUMULATED PAYMENTS ($J = 11$).

	Separate Modelling			CIP Modelling	
Mean	Function	Estimation	t-ratio	Estimation	t-ratio
	Intercept	−5.19	5.61	−3.51	4.59
	j	−0.24	3.34	−0.15	3.49
	$\log(j)$	−7.19	8.68	−5.25	5.97
	$\log(1 + j)$	10.10	6.92	7.17	5.38
Variance	Intercept	0.01	1.15	0.009	6.22
	m^2	0.06	6.21	0.09	2.30

TABLE 2

INCURRED CLAIMS ($J = 4$).

	Separate Modelling			CIP Modelling	
Mean	Function	Estimation	t-ratio	Estimation	t-ratio
	Intercept	−0.23	3.14	−0.24	4.41
	i	0.08	4.72	0.06	5.97
	i^2	−0.005	4.09	−0.004	5.39
	j	0.24	3.84	0.24	4.82
	$\log(j)$	−0.63	4.76	−0.56	5.32
Variance	Intercept	0.004	2.16	0.002	6.79
	m^2	0.32	4.21	0.784	3.22
rho		0.28		0.26	3.41

The best set of basic functions and the optimal thresholds are selected according to the Akaike's criterion.

The basic functions retained, the estimation of their coefficients, the t-ratio statistics, and J are given in Tables 1 and 2. The variances, i.e. the diagonal terms of $\Omega(i, j, \theta)$, are assumed to be affine functions of the square of the corresponding mean. The estimation of the coefficients of this affine functions and the associated t-ratios are also given in Tables 1 and 2.

It is seen that all these estimations are highly significant. They will be used as starting values for the (pseudo) ML method described above for the estimation of the CIP model. In this second stage, the correlation function $\rho(i, j)$ appearing in $\Omega(i, j)$, as well as the constraints $X_{1,i,n} = X_{2,i,n}$ have been taken into account. Different specifications for $\rho(i, j)$ have been tested and a constant function has been retained. The estimations of the parameters and the corresponding t-ratio

TABLE 3

VALUES AT THE ULTIMATE DEVELOPMENT YEAR (10^3 EUROS).

	Chain Ladder Payment	Chain Ladder Incurred	Munich–Re	PIC without Dependence	PIC with Dependence	CIP
1997	5,909	7,177	7,177	7,189	7,177	7,026
1998	3,698	4,711	4,711	4,499	4,484	4,588
1999	5,688	9,002	9,038	6,924	6,892	8,793
2000	5,082	7,040	7,046	6,185	6,157	6,960
2001	8,803	9,167	8,908	10,727	8,419	9,527
2002	6,662	6,976	6,818	8,127	6,204	7,146
2003	7,344	11,226	11,281	9,003	10,084	10,755
2004	5,548	4,427	4,114	6,811	3,815	4,565
2005	5,842	4,611	4,123	7,178	4,010	4,795
2006	7,039	18,770	20,812	8,657	17,085	16,839
2007	3,441	3,847	3,631	4,292	3,355	3,782
2008	3,428	3,980	3,756	4,341	3,416	3,600
2009	3,365	4,178	4,070	4,316	3,599	3,384
2010	2,022	2,836	2,916	3,142	2,449	1,704
Total	73,877	97,954	98,405	91,391	87,146	93,471

of the CIP model are also given in Tables 1 and 2. It is interesting to see that these estimations are, in general, rather different from the initial values and this shows the importance of jointly taking into account the information contained in the cumulated payments and the incurred claims. It is also worth noting that all the coefficients are highly statistically significant. As mentioned in Section 3.4, the CIP method also allows to propose an estimation for the ultimate development year N and we check that when some development profiles are highly volatile, like in the data considered here, the estimation of N may be large. In our case, we find $N = 31$.

5.2. Values at the ultimate development year

In the CIP model, the predicted values of the cumulated payments and of the incurred claims at the ultimate development date are, by construction, the same. It is interesting to compare these estimated ultimate values with the ones provided by the Chain Ladder method applied to the cumulated payments, by the Chain Ladder method applied to the incurred claims, by the Munich–Re method, by the PIC method without dependence and with tail development factor (Merz and Wüthrich, 2010) and by the PIC methods with dependence and without tail factor (Happ and Wüthrich, 2013). These values are displayed in Table 3. The two Chain Ladder provide very different results, the total over the accident years being $73,877 \times 10^3$ for the cumulated payments and $97,954 \times 10^3$ for the incurred claims. The results of the Munich–Re method are similar to

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

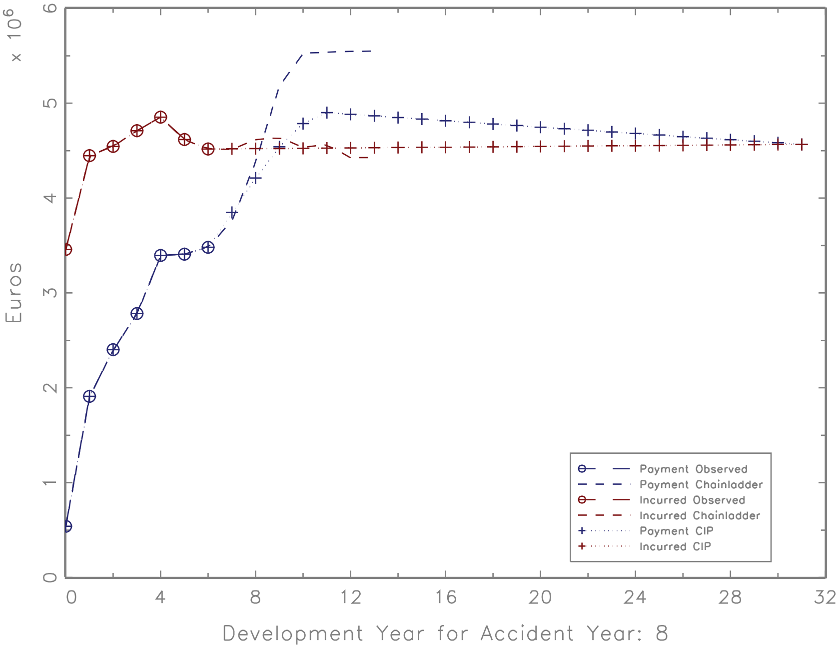

FIGURE 1: Prediction of the incurred and paid claims: year 2004. (Color online)

those of the Chain Ladder method for incurred claims. The CIP method provides, in general, values which are between the two Chain Ladders. In particular, the total is $93,471 \times 10^3$. As far as the two versions of the PIC methods are concerned, we see that the PIC method without dependence has difficulties in capturing extreme information like the very high value (18,773) of the incurred claims of accident year 2006 and development year 2010, and, therefore, provides a value at the ultimate development year (8,657) which seems to be underestimated. The PIC method with dependence captures this extreme information very differently (17,085 for the ultimate value of the accident year 2006) and in a way which is similar to that of our CIP method (16,839). However, the sum over the accident years of all the ultimate values is larger for our CIP method than for two PIC methods, apparently because of the lack of dependence in the first one and the lack of tail development factor for the second one.

Figure 1 (resp. 2) shows the predictions of the cumulated payments and of the incurred claims provided by the Chain Ladder and the CIP methods for accident year 2004 (resp. 2008). In both cases, the Chain Ladder method provides very different values for the two variables at the largest observed development horizon, i.e. 14, and the ultimate common value proposed by the CIP method is between these two values.

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

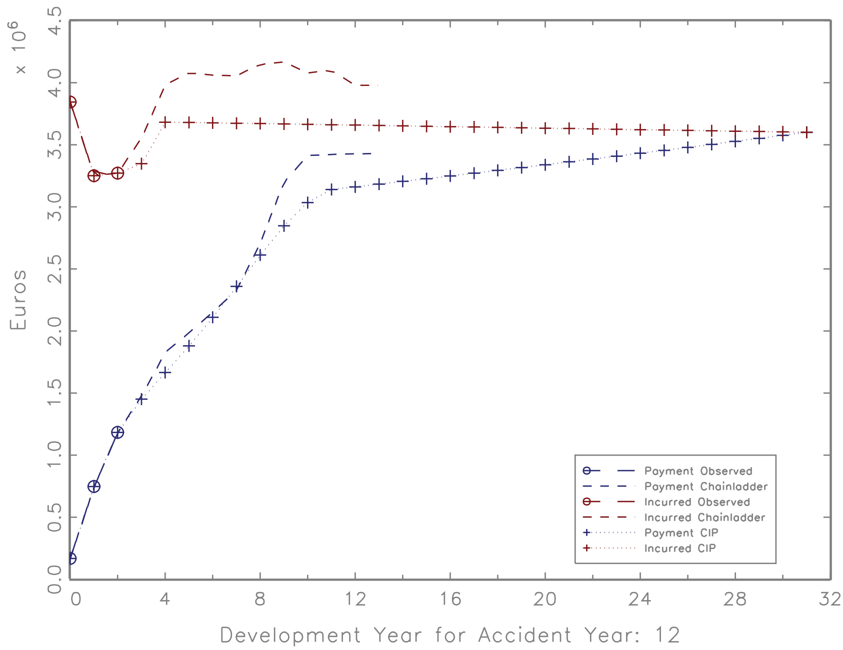

FIGURE 2: Prediction of the incurred and paid claims: year 2008. (Color online)

Figures 3 and 4 provides the whole prediction surfaces of the cumulated payments and of the incurred claims. By construction, the profiles at the ultimate development horizon are identical.

5.3. Values at Risk of the CDR's

In a previous study only based on incurred claims (see Koenig *et al.* (2015)), we have stressed the importance of two elements in the computation of the VaRs of CDR namely, the non-Gaussianity of the distributions and the updating of the estimations. As we shall see, the importance of these features are strongly confirmed by the CIP method.

First, let us test the Gaussianity of the components of the normalized vectors ε_i defined in Equation (15) and estimated by $\hat{\varepsilon}_i$ defined in Equation (17). Since $n = 14$, the Jarque–Bera statistic, given in (18), becomes

$$91\left(\frac{S^2}{3}+\frac{(K-3)^2}{12}\right),$$

where S and K are, respectively, the empirical skewness and kurtosis of the $\hat{\varepsilon}_{i,k}$.

If the errors are Gaussian, the Jarque–Bera statistic is asymptotically distributed as $\chi^2(2)$ and the null hypothesis of Gaussianity should be rejected if

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

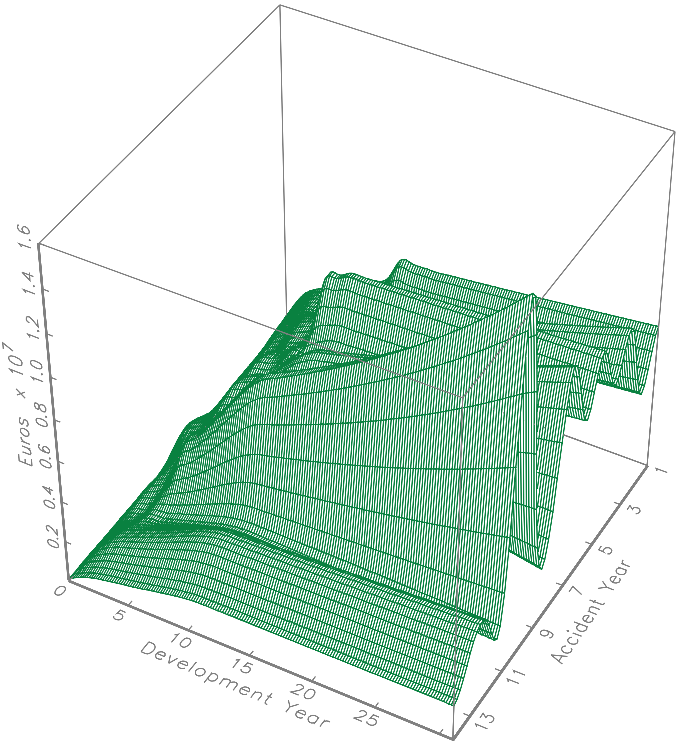

FIGURE 3: Prediction surface of the cumulated payments. (Color online)

the numerical value of this statistic is larger than the critical values, which are 4.6, 6.0, 9.2 for the 10%, 5% and 1% levels, respectively. Since the value found is 80.5, the normality assumption is very strongly rejected. This non-normality is confirmed by Figure 5 showing the kernel-based estimation of the density of the ε_i compared with the standard Gaussian density: a much ticker right tail is observed.

If follows that the appropriate computation of the VaR's of the CDR's should not assume Gaussianity and therefore should be based on the method described in Section 4.4. Moreover, it is important to measure the impact of the updating of the estimations of the parameters. Table 4 gives the results.

Let us consider the global 99.5% VaR. Wrongly assuming Gaussianity leads to a VaR equal to $3,779 \times 10^3$ instead of $4,954 \times 10^3$ when there is no updating and a VaR equal to $5,031 \times 10^3$ instead of $6,695 \times 10^3$ when there is updating. The price to pay for wrongly assuming Gaussianity is very high: an under-estimation of approximately 25%.

The price to pay for omitting updating is of the same order of magnitude. It moves from $3,779 \times 10^3$ to $5,031 \times 10^3$ in the Gaussian case and from $4,954 \times 10^3$

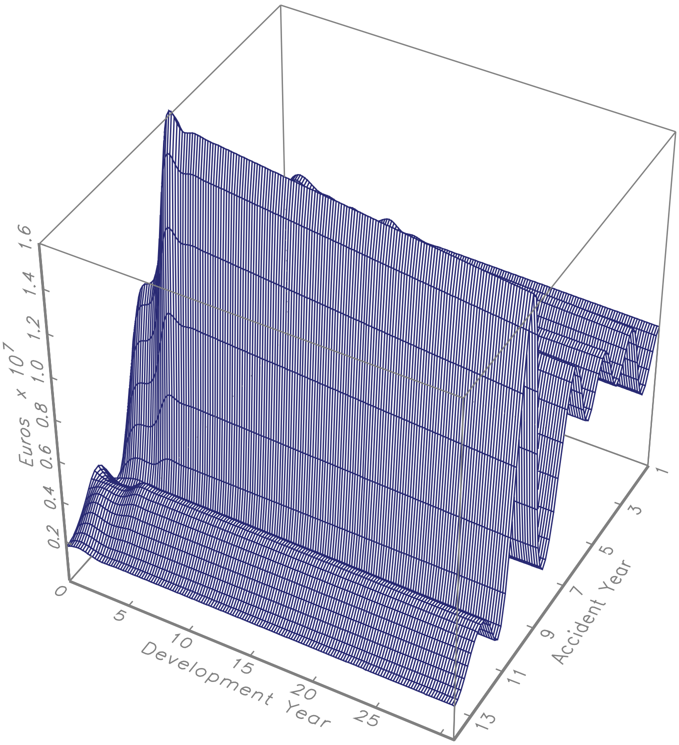

FIGURE 4: Prediction surface of the incurred claims. (Color online)

to $6,695 \times 10^3$ in the non-Gaussian case. Cumulating both mistakes leads to an under-estimation of approximately 44%.

6. CONCLUDING REMARKS

We proposed a flexible statistical modelling, called the CIP method, allowing to take into account simultaneously the payments and the incurred claims in the prediction of future claims. This method is semi-parametric since it does not assume a precise shape of the distributions but only concentrates on the first two moments. In particular, Gaussianity of the growth rates, i.e. log-Gaussianity of the levels, is not assumed and is in fact strongly rejected in our application. Moreover, our CIP method also allows to estimate correlations, the ultimate development year, the CDR's and their VaR's which are measures of reserve risk recommended by the regulatory authorities. The techniques derived in this paper could be extended in several directions. In particular, it would be interesting to derive a CIP method treating simultaneously several business lines. This kind of development is left for future research.

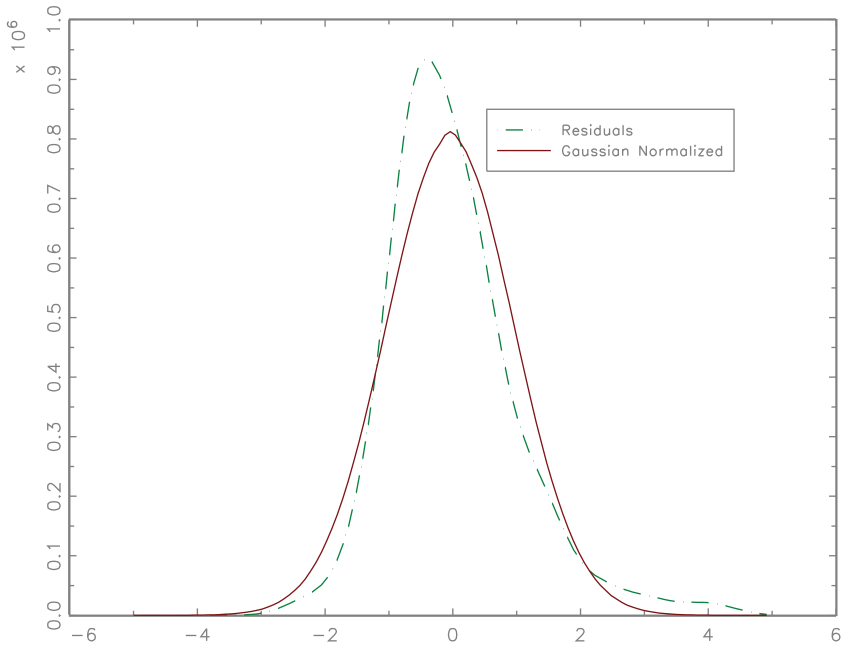

FIGURE 5: Estimated density function of the residuals. (Color online)

TABLE 4

VAR'S OF THE CDR'S (10^3 EUROS).

	Without Updating		With Updating	
	Gaussian	Non-Gaussian	Gaussian	Non-Gaussian
1997	914	1,252	901	1,281
1998	583	827	555	781
1999	1,118	1,610	1,156	1,606
2000	895	1,277	916	1,181
2001	1,251	1,751	1,222	1,669
2002	919	1,305	942	1,352
2003	1,409	1,929	1,457	2,096
2004	581	848	611	866
2005	619	894	649	871
2006	2,201	3,108	2,451	3,316
2007	1,459	2,187	1,263	1,784
2008	531	765	604	885
2009	490	709	749	1,105
2010	419	612	668	1,001
Sum of VaR's	13,394	19,081	14,150	19,803
Global VaR	**3,779**	**4,954**	**5,031**	**6,695**

REFERENCES

BOOR, J. (2006) Estimating tail development factors: What to do when the triangle runs out. *CAS Forum Winter*, 345–390.

CHOIRAT, C. and SERI, R. (2012) Discrete parameter models. *Statistical Science*, **27**, 278–293.

DAHMS, R. (2008) A loss reserving methods for incomplete data. *Bulletin Swiss Association of Actuaries*, **1&2**, 127–148.

DAHMS, R., MERZ, M. and WÜTHRICH, M. (2009) Claims development result for combined claims incurred and claims paid data. *Bulletin Français d'Actuariat*, **9**(18), 5–39.

GOURIEROUX, C. and MONFORT, A. (1996) *Statistics and Econometric Models*, vol. 2. New York: Cambridge University Press.

GOURIEROUX, C., MONFORT, A. and TROGNON. (1984) Pseudo maximum likelihood methods: Theory. *Econometrica*, **52**, 680–700.

HALLIWELL, L.J. (1997) Conjoint prediction of paid and incurred losses. *CAS Forum*, 241–379.

HALLIWELL, L.J. (2009) Modeling paid and incurred losses together. *CAS Forum*, 1–50.

HAPP, S., MERZ, M. and WÜTHRICH, M.V. (2012) Claims development result in the paid incurred claim reserving method. *Insurance, Mathematics and Economics*, **51**(1), 66–72.

HAPP, S. and WÜTHRICH, M.V. (2013) Paid-incurred chain reserving with dependence modelling. *ASTIN Bulletin*, **43**(1), 1–20.

KOENIG, E., LE MOINE, P., MONFORT, A. and RATIARISON, E. (2015) Evaluating reserve risk in a regulatory perspective. *Journal of Insurance Issues*, **38**(2), 151–183.

LANCASTER, T. (2000) The incidental parameter problem since 1978. *Journal of Econometrics*, **95**, 391–413.

LIU, H. and VERRALL, R. (2010) Bootstrap estimation of the predictive distributions of reserves using paid and incurred claims. *Variance*, **4**(2), 121–135.

MERZ, M. and WÜTHRICH, M.V. (2010) Paid-Incurred chain claims reserving method. *Insurance Mathematics and Economics*, **46**, 568–579.

MERZ, M. and WÜTHRICH, M.V. (2013) Estimation of tail development factors in the Paid-Incurred chain reserving method. *Variance*, **7**(1), 61–73.

MOREIRA, M. (2009) A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics*, **37**, 3660–3696.

NEYMAN, J. and SCOTT, E.L. (1948) Consistent estimates based on partially consistent observations. *Econometrica*, **59**, 347–378.

PETERS, G., DONG, A. and KOHN, R. (2014) A copula based Bayesian approach for paid-incurred claims models for non-life insurance reserving. *Insurance Mathematics and Economics*, **30**(59), 258–278.

PETERS, G., WÜTHRICH, M. and SHEVCHENKO, P. (2010) Chain ladder method: Bayesian bootstrap versus classical bootstrap. *Insurance Mathematics and Economics*, **47**, 36–51.

PIGEON, M., ANTONIO, K. and DENUIT, M. (2014) Individual loss reserving using paid-incurred data. *Insurance Mathematics and Economics*, **58**, 121–131.

POSTHUMA, B., CATOR, E.A., VEERKAMP, W. and VAN ZWET, E.W. (2008) Combined analysis of paid and incurred losses. *Casualty Actuarial Society E-Forum*, 272–293.

QUARG, C. and MACK, T. (2004) Munich chain ladder. *Blatter DGVFM Band,* XXVI, 597–630.

VENTER, G.G. (2008) Distribution and value of reserves using paid and incurred triangle. CAS Forum Fall.

WÜTHRICH, M.V. and MERZ, M. (2013) *Financial Modeling, Actuarial Valuation and Solvency in Insurance*. Berlin, Heidelberg: Springer.

GILLES DUPIN
Groupe MONCEAU, Paris
E-Mail: gdupin@monceauassurances.com

EMMANUEL KOENIG
Groupe MONCEAU, Paris
E-Mail: ekoenig@monceauassurances.com

PIERRE LE MOINE
Groupe MONCEAU, Paris
E-Mail: plemoine@monceauassurances.com

ALAIN MONFORT (Corresponding author)
CREST Paris
E-Mail: monfort@ensae.fr

ERIC RATIARISON
Groupe MONCEAU, Paris
E-Mail: eratiarison@monceauassurances.com

APPENDIX A

Proof of Proposition 3.1

Let us first consider the joint distribution of $\begin{pmatrix} \tilde{Y}_i \\ d' X_{i,N} \end{pmatrix}$. Since $X_{i,N} = X_{i,0} + \sum_{j=1}^{N} Y_{i,j}$, this joint distribution (given $X_{i,0}$) is Gaussian.

Its mean is

$$\begin{pmatrix} \tilde{m}_i \\ d' X_{i,0} + d' \sum_{j=1}^{N} m_{i,j} \end{pmatrix} = \begin{pmatrix} \tilde{m}_i \\ a_i \end{pmatrix}$$

and its variance–covariance matrix is

$$\begin{pmatrix} \Omega_{i,1} \dots & 0 & \tilde{c}_i \\ 0 & \Omega_{i,N} & \\ \tilde{c}_i' & & b_i \end{pmatrix} = \begin{pmatrix} \tilde{\Omega}_i & \tilde{c}_i \\ \tilde{c}_i & b_i \end{pmatrix}$$

with

$$\begin{aligned} b_i &= V(d' X_{i,N}) = \sum_{j=1}^{N} d' \Omega_{i,j} d, \\ \tilde{c}_i &= \text{cov}(\tilde{Y}_i, d' X_{i,N}), \\ &= \begin{pmatrix} c_{i,1} \\ \vdots \\ c_{i,N} \end{pmatrix}, \end{aligned}$$

and $c_{ij} = \Omega_{i,j} d$ and therefore $\tilde{c}_i = \tilde{\Omega}_i F_N d$, with $F_N = (I_2, \dots, I_2)'$.

Applying a standard formula for conditional Gaussian distributions [see Gourieroux and Monfort (1996) Appendix 3.3], we see that the conditional distribution of $\tilde{Y}_i$ given $d' X_{i,N} = 0$ is

$$N\left(\tilde{m}_i - \frac{\tilde{c}_i a_i}{b_i}, \tilde{\Omega}_i - \frac{\tilde{c}_i \tilde{c}_i}{b_i} \right). \qquad \square$$

APPENDIX B

Proof of Proposition 3.3

Lemma

Let β a vector such that $\| \beta \| \neq 1$ (with $\| \beta \|^2 = \beta'\beta$), the matrix $I - \beta\beta'$ is invertible and

$$(I - \beta\beta')^{-1} = I + \frac{\beta\beta'}{1 - \| \beta \|^2}.$$

Proof:

$$\begin{aligned} (I - \beta\beta')\left(I + \frac{\beta\beta'}{1 - \| \beta \|^2} \right) &= I - \beta\beta' + \frac{\beta\beta'}{1 - \| \beta \|^2} - \frac{\beta\beta' \| \beta \|^2}{1 - \| \beta \|^2} \\ &= I. \end{aligned} \qquad \square$$

Let us now consider the matrix

$$\Omega_i - \frac{c_i c_i'}{b_i} = \Omega_i^{1/2} \left(I - \frac{\Omega_i^{-1/2} c_i}{b_i^{1/2}} \frac{c_i' \Omega_i^{-1/2}}{b_i^{1/2}} \right) \Omega_i^{1/2}$$

setting $\beta_i = \dfrac{\Omega_i^{-1/2} c_i}{b_i^{1/2}}$, we get

$$\Omega_i - \frac{c_i c_i'}{b_i} = \Omega_i^{1/2} (1 - \beta_i \beta_i') \Omega^{1/2},$$

and applying the lemma we get

$$\begin{aligned} \left(\Omega_i - \frac{c_i c_i}{b_i} \right)^{-1} &= \Omega_i^{-1/2} \left(I + \frac{\beta_i \beta_i'}{1 - \| \beta_i \|^2} \right) \Omega_i^{-1/2}, \\ &= \Omega_i^{-1} + \frac{\Omega_i^{-1} c_i c_i' \Omega_i^{-1}}{b_i \left(1 - \dfrac{c_i' \Omega_i^{-1} c_i}{b_i} \right)} \\ &= \Omega_i^{-1} + \frac{\Omega_i^{-1} c_i c_i' \Omega_i^{-1}}{b_i - c_i' \Omega_i^{-1} c_i}. \end{aligned} \qquad \square$$

https://doi.org/10.1017/asb.2017.36 Published online by Cambridge University Press

APPENDIX C

Payments (10^3 euros)

455	1,941	2,647	3,006	3,248	3,608	3,855	4,215	4,427	4,811	5,890	5,910	5,907	5,910
659	1,780	2,410	2,954	3,103	3,192	3,302	3,596	3,670	3,678	3,678	3,680	3,697	
480	1,536	2,340	2,831	3,343	3,613	3,920	4,815	5,499	5,643	5,664	5,678		
183	1,433	2,039	2,449	2,985	3,195	3,637	3,795	4,597	4,913	5,061			
131	1,372	2,081	3,001	3,289	3,537	4,105	4,211	4,874	8,227				
139	1,266	1,975	2,385	2,647	3,177	3,508	3,691	5,267					
372	1,371	2,067	2,614	3,869	4,467	4,774	4,984						
539	1,909	2,401	2,781	3,394	3,409	3,482							
212	1,042	1,889	2,396	3,204	3,381								
203	1,098	1,615	2,184	3,751									
209	804	991	1,486										
168	747	1,183											
161	801												
110													

Incurred claims (10^3 euros)

7,430	8,289	8,223	7,849	7,786	7,561	7,623	7,284	7,400	7,696	7,317	7,318	7,177	7,177
6,654	7,319	6,453	6,340	6,318	6,014	5,375	5,414	5,385	5,078	5,013	4,933	4,711	
6,555	7,795	7,394	7,990	7,540	8,596	8,314	8,713	8,931	8,995	9,072	9,277		
5,804	6,573	6,248	6,187	6,462	6,308	7,238	7,104	7,301	7,481	7,213			
5,527	6,644	7,551	9,198	9,008	9,423	9,732	9,324	9,561	9,601				
4,066	5,945	6,537	6,332	6,272	6,797	6,651	7,077	7,272					
3,301	5,169	5,808	8,083	11,294	11,922	11,490	11,441						
3,459	4,446	4,544	4,709	4,853	4,617	4,517							
2,376	4,073	4,065	4,337	4,817	4,728								
6,035	10,398	11,191	13,269	18,773									
2,470	3,260	3,452	3,439										
3,844	3,251	3,271											
2,733	3,362												
1,797													