



UNIVERSITÀ DEGLI STUDI
DI PERUGIA

ISSN 2385-2275

**Working papers of the
Department of Economics
University of Perugia (IT)**

**On the Estimation of the
Undertaking-Specific
Parameters and the
Related Hypothesis
Testing**

Updated version

**Massimo De Felice
Franco Moriconi**

Working paper No. 25

October 2016

On the Estimation of the Undertaking-Specific Parameters and the Related Hypothesis Testing

Updated version

Massimo De Felice – Sapienza, Università di Roma

Franco Moriconi – Università di Perugia

Contents

I The theoretical models and the related hypothesis testing	3
1. The theoretical models underlying the standardised methods for USP	4
1.a. Reference Model for Method 1 (Model M1)	4
1.b. Reference model for Method 2 (Model M2)	10
2. Testing the hypotheses of the theoretical models	12
2.a. Hypothesis testing for Model M1	12
2.a.1. Hypothesis on the mean	12
2.a.2. Hypothesis on the variance	14
2.a.3. An example of a market-wide test of M1V hypothesis	15
2.a.4. Hypothesis on the distribution	17
2.a.5. Comparing the testing methods	21
2.a.6. Appropriateness of maximum likelihood method . . .	25
2.b. Hypothesis testing for Model M2	28
2.b.1. Hypothesis on the conditional mean and variance . . .	29
2.b.2. Independence hypothesis. Test on time series residuals	30
2.b.3. Independence hypothesis. Test on Pearson residuals .	33

This paper is the English version of *Sulla stima degli Undertaking Specific Parameters e la verifica delle ipotesi. Versione aggiornata*. Working paper No. 16, Department of Economics, University of Perugia, November 2015.

II	Application to entity-specific data	35
3.	Premium Risk – Model M1	35
3.a.	Specification of the input data	35
3.b.	Application of the method	36
3.c.	On the minimisation technique	37
4.	Reserve Risk – Model M1	37
4.a.	Specification of the input data	37
4.b.	Application of the method	38
4.c.	On the minimisation technique	39
5.	Reserve Risk – Model M2	39
5.a.	Specification of the input data	39
5.b.	Application of the method	40
	Appendix	40
	A. Autocorrelation and heteroscedasticity	40
	B. Resampling of individual data	41

Introductory note – This working paper illustrates the calculation process for estimating the undertaking-specific parameters (USP) as defined in Solvency II, taking into account the underlying theoretical basis. The USPs considered here are the unit standard deviation for the premium risk and reserve risk submodules of non-life insurance; the analysis does not take into account the entity-specific adjustment factors for non-proportional reinsurance.

For each calculation method of the unit standard deviation we present the formal settings proposed by European Commission; the theoretical principles are recalled, the appropriate methods for hypothesis testing and for assessing the “goodness-of-fit” to data are described, data necessary for calculation is specified and the relevant computational issues are discussed.

This working paper has practical motivations; it coordinates “useful documentation” on methodologies, criteria, algorithms, types of analysis for the “determination of the specific parameters” (as required by IVASS in [15] point (g)). In order to facilitate immediate application, we occasionally recall notions deemed as standard in best practice, for instance on hypothesis testing.

Changes in this version – A previous version of this paper has been published in April 2015, as Working Paper No. 9 of the Department of Economics of University of Perugia. Compared to the previous version, this paper extends the theoretical and methodological analysis to some new topics about USP that have been discussed in recent months. Furthermore it specifies some practical approaches that seem to have become part of the current best practice.

In summary, in Section 1.a. it has been widened and deepened the analysis of the theoretical model underlying “Method 1”; in Part II some issues concerning input data, in particular that net of reinsurance, have been clarified and updated. Furthermore two short appendices have been added: in Appendix A there are some considerations on hypothesis testing when autocorrelation and heteroscedasticity are present in the data; in Appendix B the resampling of individual data by Block Bootstrap methods is considered.

The authors are grateful to Stefano Cavastracci for the useful discussions that allowed the clarification of some critical issues contained in this updated version.

Part I

The theoretical models and the related hypothesis testing

1 The theoretical models underlying the standardised methods for USP

Delegated Acts specify two *standardised methods* to calculate the undertaking-specific unit standard deviations. *Method 1* can be used for both the premium risk and the reserve risk submodule, *Method 2* is an alternative approach which can be applied only to the reserve risk submodule. Each method is based on a specific underlying stochastic model, that we briefly describe as follows.

1.a Reference Model for Method 1 (Model M1)

The reference model for the unit standard deviation according to Method 1 has been defined by the Joint Working Group on Non-Life and Health NSLT Calibration (JWG) in [9], in the context of the market-wide calibration study for the premium and reserve risk factors in the underwriting risk module of the SCR standard formula.

The theoretical model underlying Method 1 (Model M1 hereafter), is one of the four alternative models analyzed and tested by the JWG in the calibration activity on European market data. For each segment of the non-life activity, these models consider a random variable Y , the variance of which must be determined and estimated based on its theoretical relations with an explanatory variable X , taken as a volume measure. In the applications to premium risk, the dependent variable Y corresponds to the aggregated claims cost of a given accident year and the independent variable X represents the corresponding level of the earned premiums. In the applications to reserve risk, the two variables X and Y represent the ultimate cost estimated respectively at the beginning and at the end of the reference year for claims occurred in the previous years.

Model M1 underlying USP calculation seems to be the one characterised in the JWC calibration study as the class of “Lognormal Models, Second Variance Parametrisation”. The model is based on the following assumptions.

M1M - *Assumption on the mean*:

$$\mathbf{E}(Y) = \beta X .$$

M1V - *Assumption on the variance:*

$$\mathbf{Var}(Y) = \beta^2 \sigma^2 \left[(1 - \delta) \bar{X} X + \delta X^2 \right], \quad (1.1)$$

where:

$$\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t,$$

is the sample mean of a yearly time series X_1, X_2, \dots, X_T of observations of X .

M1D - *Assumption on the distribution:*

$$\ln Y \sim \text{Normal}(\mu, \omega),$$

where:

$$\omega = \ln \left\{ 1 + \sigma^2 \left[(1 - \delta) \bar{X} / X + \delta \right] \right\}, \quad \mu = \ln(\beta X) - \frac{\omega}{2}. \quad (1.2)$$

In order to estimate the model we need to estimate the parameters β, σ and δ . In particular,

- $\delta \in [0, 1]$ is a *mixing parameter*. If $\delta = 1$ the variance of Y has a quadratic relation with X , while if $\delta = 0$, the variance of Y is proportional to X .
- σ approximates (in practice, it coincides with) the variation coefficient of Y , $\mathbf{Cv}(Y) = \mathbf{Std}(Y)/\mathbf{E}(Y)$. Therefore an estimate of σ provides the value of the undertaking-specific unit standard deviation for premium risk or reserve risk (depending on how the random variables X and Y are interpreted).

Remark. Assumption M1V that specifies the variance of Y as a quadratic function of X , is motivated by the JWG as being a “realistic” extension of the Compound Poisson model often used for the underwriting risk within the actuarial practice (see e.g. [24], Chapter 3). In the Compound Poisson model, whose parameters are constant over time, the mean and the variance of the aggregated claims cost is a linear function of portfolio size. If we move from an assumption of constant parameters to one of time-varying parameters according to a stochastic (stationary) process, we still obtain a linear expression for the mean but an expression of the type $\mathbf{Var}(Y) = \sigma_1^2 \bar{X} X + \sigma_2^2 X^2$ for the variance. If we assume in addition that $\mathbf{Var}(Y)$ is proportional to β^2 (this is “the second variance parametrisation”) we obtain, with some manipulations, expression (1.1). This result implies that the variation coefficient of Y is independent of β . Moreover it allows to obtain maximum likelihood parameter estimates without using too complex optimisation procedures. ■

Remark. It should be pointed out that, within the JWG calibration study, the constant \bar{X} is defined as the arithmetic mean of the observations of X taken on all the companies operating in the reference market. Denoting by $X_{t,i}$ the observation t of company i , in [9] one finds:

$$\bar{X} = \frac{\sum_{i=1}^N \sum_{t=1}^{T_i} X_{t,i}}{\sum_{i=1}^N T_i},$$

where N is the number of companies operating in the market (within the specified segment) and T_i is the number of available observations for company i . This factor has been introduced in the variance expression in order to make the coefficient $\beta^2 \sigma^2 (1 - \delta)$ independent of the monetary dimension. In the transposition of the JWG model from a market-wide point of view to a single-company point of view, the quantity \bar{X} has been re-defined, by the Delegated Acts, as an individual mean. This choice can further increase the model instability for time series with a short number of observations. Also, it brings explicit effects on the statistical tests of the M1V hypothesis (see Section 2.a.2). ■

Some details on the structure of Model M1

It could be useful to recall with some details the basic structure of Model M1, since JWG's document simply provides a unified presentation of the entire set of the models considered for the calibration. Moreover, compared to the original one, the model presented in the Delegated Acts contains a reparametrisation of the estimation function.

The parameter estimation of Model M1 is obtained by the maximum likelihood method applied to an undertaking-specific time series of observations:

$$(\mathbf{X}, \mathbf{Y}) = \{(X_t, Y_t); t = 1, 2, \dots, T\}.$$

These observations must be considered as independent realizations of the two-dimensional random variable (X, Y) . Let us denote by $\pi = \omega^{-1}$ the *precision* (reciprocal of the variance). If one observes that the random variable:

$$u = \ln \frac{Y}{X} + \frac{1}{2\pi} - \ln \beta,$$

is normally distributed with zero mean and variance $1/\pi$, it can be easily shown that the maximisation of the likelihood of Y is equivalent to the minimisation with respect to β, σ, δ , given (\mathbf{X}, \mathbf{Y}) , of the *loss function* (*criterion function*):

$$\ell(\beta, \sigma, \delta) = \sum_{t=1}^T \pi_t u_t^2 - \sum_{t=1}^T \ln \pi_t, \quad (1.3)$$

where, for $t = 1, 2, \dots, T$:

$$u_t = \ln \frac{Y_t}{X_t} + \frac{1}{2\pi_t} - \ln \beta, \quad (1.4)$$

and:

$$\pi_t = \frac{1}{\ln \{1 + \sigma^2 [(1 - \delta)\bar{X}/X_t + \delta]\}}. \quad (1.5)$$

This expression for the precisions π_t is obtained by the first expression in (1.2), which in turn is a consequence of the assumption M1V on the variance, that is expression (1.1), which is the “second variance parametrisation” considered by the JWG. This expression of π_t depends on σ and δ but is independent of β ¹, then (1.3) can be minimised with respect to $\ln \beta$. One obtains:

$$\ln \hat{\beta} = \frac{\sum_{t=1}^T a_t \pi_t}{\sum_{t=1}^T \pi_t},$$

with $a_t := \ln(Y_t/X_t) + 1/(2\pi_t)$. That is:

$$\ln \hat{\beta} = \frac{T/2 + \sum_{t=1}^T \pi_t \ln(Y_t/X_t)}{\sum_{t=1}^T \pi_t}. \quad (1.6)$$

Using this expression (which also depends only on σ and δ) the minimisation of the criterion function can be reduced to a two-variables problem, consisting in the minimisation of:

$$\ell(\delta, \sigma) = \sum_{t=1}^T \pi_t \left(\ln \frac{Y_t}{X_t} + \frac{1}{2\pi_t} - \ln \hat{\beta} \right)^2 - \sum_{t=1}^T \ln \pi_t. \quad (1.7)$$

In the Delegated Acts this problem is reparametrised by replacing σ by the parameter $\gamma = \ln \sigma$, hence (1.5) is rewritten as:

$$\pi_t(\delta, \gamma) = \frac{1}{\ln \{1 + [(1 - \delta)\bar{X}/X_t + \delta] e^{2\gamma}\}}, \quad (1.8)$$

(where the functional dependence by the parameters has been explicitly indicated). Moreover the new function is introduced:

$$\hat{\sigma}(\delta, \gamma) := \sigma \hat{\beta} = e^\gamma \hat{\beta} = \exp \left[\gamma + \frac{T/2 + \sum_{t=1}^T \pi_t(\delta, \gamma) \ln(Y_t/X_t)}{\sum_{t=1}^T \pi_t(\delta, \gamma)} \right], \quad (1.9)$$

which leads to the expression:

$$\ln \hat{\beta} = -\gamma + \ln[\hat{\sigma}(\delta, \gamma)].$$

¹This independence property directly derives from the fact that, by (1.1), the coefficient of variation:

$$\mathbf{Cv}(Y) = \sigma [(1 - \delta)\bar{X}/X + \delta]^{1/2},$$

is independent of β .

Therefore the criterion function (1.7) takes the form:

$$\ell(\delta, \gamma) = \sum_{t=1}^T \pi_t(\delta, \gamma) \left\{ \ln \frac{Y_t}{X_t} + \frac{1}{2\pi_t(\delta, \gamma)} + \gamma - \ln [\hat{\sigma}(\delta, \gamma)] \right\}^2 - \sum_{t=1}^T \ln [\pi_t(\delta, \gamma)], \quad (1.10)$$

which is the expression actually provided by the official documents.

This function has to be minimised in the interval $D = \{\delta \in [0, 1], \gamma \in \mathbb{R}\}$ using an appropriate numerical optimisation procedure. The values $\hat{\delta}$ and $\hat{\gamma}$ thus obtained are the parameter estimates which provide, through expression (1.9), the maximum likelihood estimate $\hat{\sigma}(\hat{\delta}, \hat{\gamma})$ for the undertaking-specific unit standard deviation for the segment considered.

Remark. Among the available estimation methods the maximum likelihood approach has the best theoretical properties and the strongest characteristics of probabilistic consistency (at least by a Bayesian point of view). For a reliable application of the method however, the maximum likelihood point (the minimum of the loss function) must be efficiently and univocally identified. For Model M1 this is equivalent to require that the numerical procedure used for minimising the function $\ell(\delta, \gamma)$ – the form of which, obviously, depends on the data (\mathbf{X}, \mathbf{Y}) – has suitable convergence properties. ■

As required by the Delegated Acts, once the minimum of the loss function has been obtained, the estimate $\hat{\sigma}(\hat{\delta}, \hat{\gamma})$ shall be multiplied by the “correction factor” $\sqrt{(T+1)/(T-1)}$. After this correction the estimates shall be mixed with the standard market-wide parameter by applying the credibility factor prescribed by EIOPA, which is a function of the time length T of the time series used for the estimation.

On an alternative derivation of Model M1

One could propose an alternative derivation of expression (1.10) obtained by a different formulation of the MV1 assumption. Instead of (1.1), this different formulation could be given by:

$$\mathbf{Var}(Y) = \sigma^2 [(1 - \delta)\overline{X}X + \delta X^2]. \quad (1.11)$$

This alternative specification of the basic assumption would be motivated by the fact that (1.11) coincides with the “first variance parametrisation” considered by the JWG.

Under this assumption the loss function $\ell(\beta, \sigma, \delta)$ is still given by (1.3) but the precisions have the form:

$$\pi_t = \frac{1}{\ln \{1 + (\sigma^2/\beta^2) [(1 - \delta)\overline{X}/X_t + \delta]\}}, \quad (1.12)$$

which is no longer independent of β . This poses problems of mathematical/computational tractability in the minimisation of the $\ell(\beta, \sigma, \delta)$ function². Trying to overcome these difficulties one could adopt a “pragmatic” approach consisting in defining the new parameter:

$$\gamma := \ln \frac{\sigma}{\beta}. \quad (1.13)$$

With this choice the loss function $\ell(\beta, \sigma, \delta)$ becomes a function of γ, σ and δ , and the precision π_t depends now only on γ and δ . The tractability of the problem of minimising $\ell(\gamma, \sigma, \delta)$ is then recovered. In fact since π_t is independent of σ , one can minimise $\ell(\gamma, \sigma, \delta)$ with respect to σ , obtaining:

$$\hat{\sigma}(\delta, \gamma) := \exp \left(\frac{\sum_{t=1}^T b_t \pi_t}{\sum_{t=1}^T \pi_t} \right), \quad (1.14)$$

with $b_t := \ln(Y_t/X_t) + 1/(2\pi_t) + \gamma$. Using this expression the loss function (1.3) becomes a function only of the δ and γ variables; since $\ln \beta = \ln \sigma - \gamma$, it takes the form:

$$\ell(\delta, \gamma) = \sum_{t=1}^T \pi_t \left(\ln \frac{Y_t}{X_t} + \frac{1}{2\pi_t} + \gamma - \ln \hat{\sigma} \right)^2 - \sum_{t=1}^T \ln \pi_t. \quad (1.15)$$

Obviously, under the transformation (1.13) expression (1.12) of π_t coincides with (1.8) and it is immediately proved that (1.14) coincides with (1.9), since $b_t = a_t + \gamma$; hence (1.15) coincides with (1.10). Therefore with this alternative approach one obtains the same criterion function specified in the Delegated Acts (furthermore the introduction of the parameter γ appears better motivated).

It should be noted however that in the previous procedure the reparametrisation (1.13) implies a redefinition of some basic quantities. In fact, if one introduces the definition (1.13), expression (1.11) takes the form $\mathbf{Var}(Y) = \beta^2 e^{2\gamma} [(1-\delta)\bar{X}X + \delta X^2]$, which is equivalent to reintroduce assumption (1.1) with a different notation (replacement of σ by e^γ). Ultimately, then, this alternative derivation of expression (1.10), though starting from the original assumption of the “first variance parametrisation”, solves the minimisation problem by implicitly transforming this assumption into the “second variance parametrisation”. Therefore it seems appropriate to consider (1.1) as the genuine variance assumption underlying Model M1.

² These difficulties are also indicated in JWG’s document in more places: *The first variance parametrisation is awkward from a mathematical and computational point of view.* ([9], Section 6); *This function [...] does not allow convenient reduction for optimisation.* ([9], Section 6.1). Despite this, in Section 4.1.1 it is stated that *eventually only the first [variance parameterization] has been used to derive the sigmas..*

1.b Reference model for Method 2 (Model M2)

The reference model for the calculation of the undertaking-specific unit standard deviation according to the second standardised method is a loss reserving model widely quoted in the actuarial literature, known as “Merz-Wüthrich model” [23]. Also this model has been experimented by the JWG within its market-wide calibration study for the reserve risk factors.

With the exception of an unessential change in the technical assumptions³, the Merz and Wüthrich model (here also referred to as Model M2), coincides with the well-known Distribution-Free Chain Ladder model (DFCL) proposed by Mack in 1993 [21]. The model, however, is applied under a different point of view, compared with the traditional approach. In Model M2 the *mean square error of prediction* (MSEP), rather than been considered in relation to the full run-off of the outstanding liabilities, is calculated under a one-year view, being related to the *Claims Development Result* (CDR) of the current accounting year. The transition from a long term view to a one-year view is required to make the measurement of uncertainty consistent with the prescriptions in Solvency II.

Remark. The use of a one-year point of view as the proper approach to solvency applications had been already introduced in 2003 in [5] with a different name – Year-End Expectation (YEE), instead of CDR – and referred to a different stochastic model, Over-Dispersed Poisson (ODP) instead of DFCL model. The explicit formulas for the MSEP in the YEE version for the DFCL model have also been derived in 2006 in [6]. The YEE point of view has been used in a field study based on paid losses data of the Motor Third Party Liability (MTL) Italian market produced by ISVAP in 2006 [8]; both the ODP and the DFCL model was used in this study. ■

For a given segment of the non-life activity, Model M2 considers the observed paid losses X of a “run-off triangle (trapezoid)” organised by accident year $i = 0, 1, \dots, I$ and development year $j = 0, 1, \dots, J$, with $I \geq J$. Therefore $X_{i,j}$ represents the “incremental” aggregated payments for claims occurred in year i made in development year j . The corresponding cumulative payments are:

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

Model M2 is based on the following assumptions:

M2I - Independence assumption. The cumulative payments $C_{i,j}$ of different accident years are stochastically independent.

³Instead of the Markov property (see the following M2CM assumption) in DFCL model only assumptions on the mean and the variance are used.

M2MC - *Markov assumption*. For $i = 0, 1, \dots, I$, the process $(C_{i,j})_{j \geq 0}$ is a Markov Chain:

$$\mathbf{P}(C_{i,j} \leq x | C_{i,0}, C_{i,1}, \dots, C_{i,j-1}) = \mathbf{P}(C_{i,j} \leq x | C_{i,j-1}).$$

M2M - *Conditional mean assumption*. For $1 \leq j \leq J$ there exist constants $f_j > 0$ such that for $0 \leq i \leq I$:

$$\mathbf{E}(C_{i,j} | C_{i,j-1}) = f_{j-1} C_{i,j-1}.$$

M2V - *Conditional variance assumption*. For $1 \leq j \leq J$ there exist constants $\sigma_j > 0$ such that for $0 \leq i \leq I$:

$$\mathbf{Var}(C_{i,j} | C_{i,j-1}) = \sigma_{j-1}^2 C_{i,j-1}.$$

Under these assumptions one obtains that the chain ladder estimators:

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{S_j}, \quad \text{with} \quad S_j = \sum_{i=0}^{I-j-1} C_{i,j}, \quad (1.16)$$

are unbiased estimators for f_j , $j = 0, 1, \dots, J-1$. Furthermore, the estimators:

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, \quad (1.17)$$

are unbiased estimators of σ_j^2 , $j = 0, 1, \dots, J-2$. If $I > J$ this expression also holds for $j = J-1$; otherwise σ_{J-1}^2 is estimated through extrapolation as follows:

$$\hat{\sigma}_{J-1}^2 = \min \left\{ \hat{\sigma}_{J-2}^2, \hat{\sigma}_{J-3}^2, \frac{\hat{\sigma}_{J-2}^4}{\hat{\sigma}_{J-3}^2} \right\}. \quad (1.18)$$

The estimate of the ultimate cost for the “open” accident years is obtained by projecting the cumulative payments of the last observed “diagonal” through the estimated chain ladder factors:

$$\hat{C}_{i,J} = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j, \quad i = I-J+1, I-J+2, \dots, I.$$

Using these estimators a closed-form expression for the MSEP estimate of the total one-year CDR of the open accident years is obtained. This well known expression is not reported here for brevity.

Remark. The MSEP includes both a *process variance* component, related to the uncertainty of in the cost development process, and an *estimation error* component, deriving from the uncertainty of the estimation of the

unknown parameters of the model. Despite of the independence assumption, this second component of uncertainty includes a covariance effect that reduces the diversification among accident years. This effect is taken into account in the expression of the total MSE_P. ■

Using Model M2, the estimation of the undertaking-specific unit standard deviation for reserve risk (in the given segment) is given by the ratio:

$$\widehat{\mathbf{Cv}}_{res} = \frac{\sqrt{\widehat{\text{MSE}}_P}}{\widehat{R}}, \quad (1.19)$$

where $\widehat{R} = \sum_{i=I-J+1}^I (\widehat{C}_{i,J} - C_{i,I-i})$ is the estimate of the outstanding loss liabilities (i.e. the undiscounted reserve estimate) provided by the model. As for the USPs given by Method 1, this estimate shall be mixed with the market-wide parameter prescribed by the Standard Formula (for the given segment) using the credibility factor c established by EIOPA.

2 Testing the hypotheses of the theoretical models

2.a Hypothesis testing for Model M1

Following the requirements of Delegated Acts, in order to verify that the reference model fits the entity-specific data, statistical tests have to be performed on the three assumptions of Model M1 introduced in Section 1.a:

M1M – *Hypothesis on the mean*: linear relation (proportionality) between $\mathbf{E}(Y)$ and X ,

M1V – *Hypothesis on the variance*: quadratic relation between variance $\mathbf{Var}(Y)$ and X ,

M1D – *Hypothesis on the distribution*: lognormality of Y .

One also needs to verify the:

ML – *Appropriateness of the maximum likelihood method* used for the estimation.

2.a.1 Hypothesis on the mean

In order to verify the M1M assumption of “linear proportionality” between the means of Y e X it is sufficient to perform a classical linear regression analysis between $\mathbf{E}(Y)$ and X , with or without intercept. If one assumes that the observations Y_t can be interpreted as unbiased estimates of $\mathbf{E}(Y)$, one can perform the analysis directly on the undertaking-specific time series:

$$(\mathbf{X}, \mathbf{Y}) = \{(X_t, Y_t); t = 1, 2, \dots, T\}, \quad (2.20)$$

according to the model (even with no intercept):

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, 2, \dots, T,$$

where the ε_t variables are independent error terms with zero mean and constant variance σ_ε^2 .

Furthermore, one can perform a market-wide linear regression analysis, using publicly available data on a sample of N companies similar to the undertaking which is making the estimate. In this case, data is given by:

$$\{(\mathbf{X}, \mathbf{Y})_i; i = 1, 2, \dots, N\} = \{(X_{t,i}, Y_{t,i}); i = 1, 2, \dots, N, t = 1, 2, \dots, T_i\}, \quad (2.21)$$

and one considers the model (even with no intercept):

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + \varepsilon_i, \quad i = 1, 2, \dots, N, \quad (2.22)$$

where:

$$\bar{X}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} X_{t,i}, \quad \bar{Y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} Y_{t,i},$$

are the sample means of X and Y , respectively, of company i (usual meaning of the error terms).

In order to test the M1M hypothesis one should measure the overall significance of the model by checking the value of the F statistic (which concerns the hypothesis that all parameters are zero except for the intercept) and the corresponding p -value. Moreover one could consider the level of “explained variance” of the regression by calculating the R^2 coefficient (coefficient of determination). As concerning the estimate of the single parameters, one has to verify that the coefficient β_1 is significantly different from zero and, in the case with intercept, also that β_0 is not significantly different from zero. As usual, one assumes the parameter being equal to zero as the null hypothesis and one adopts the classical hypothesis tests available for these applications. The standard approach for assessing the parameter significance is a two tailed test based on the t -Student statistic at a given significance level α (e.g. $\alpha = 10\%$). To reject the null hypothesis one will consider the p -value associated to the test statistic (in this case, the probability that the absolute value of the random variable t is higher than the observed value).

Remark. In model selection applications, it is a good practice to perform comparisons among models by using a variety of goodness-of-fit indices normalized for the number of observations and the number of parameters. One can consider, for example, the SSE (Sum of Squared Errors) adjusted by the *Squared Degree of Freedom* criterion (SDF), the SSE adjusted by the *Akaike Information Criterion* and the SSE adjusted by the *Bayesian Information*

Criterion. Since the statistical tests considered here do not require a comparison between alternative models, the use of these indexes is not necessary and we can limit the calculation to just one of these goodness-of-fit measures (for instance the SSE-SDF). ■

Additional remarks concerning autocorrelation and heteroscedasticity can be found in Appendix A.

2.a.2 Hypothesis on the variance

In order to verify the M1V hypothesis of variance $\mathbf{Var}(Y)$ being a quadratic function of X , it is convenient to use a market-wide approach, since it is generally not possible to obtain a reliable set of independent observations of $\mathbf{Var}(Y)$ using only entity-specific data (\mathbf{X}, \mathbf{Y}) . For alternative approaches based on the resampling of individual data, however, see Appendix B.

As a practical approach, let us consider a sample of market observations:

$$\{(\mathbf{X}, \mathbf{Y})_i; i = 1, 2, \dots, N\} = \{(X_{t,i}, Y_{t,i}); i = 1, 2, \dots, N, t = 1, 2, \dots, T_i\},$$

referred to a set of N companies similar to the one that is performing the estimate. In order to test the variance hypothesis one can estimate on that data the model:

$$\widehat{\mathbf{Var}}_i(Y) = \beta_0 + \beta_1 \bar{X}_i + \beta_2 \bar{X}_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, N, \quad (2.23)$$

where:

- $\widehat{\mathbf{Var}}_i(Y)$ is an estimate of the variance $\mathbf{Var}_i(Y)$ of company i ,
- \bar{X}_i is the sample mean $\sum_{t=1}^{T_i} X_{t,i}/T_i$.

Given the structure of M1V assumption, the parameters in (2.23) shall have the form:

$$\beta_0 = 0, \quad \beta_{1,i} = \beta^2 \sigma_i^2 (1 - \delta_i) \bar{X}_i, \quad \beta_{2,i} = \beta_i^2 \sigma_i^2 \delta_i,$$

where the index i denotes the dependence on the single company. In fact, under the assumptions of Model M1 both δ_i and $\sigma_i = \beta \exp \gamma_i$ are entity-specific. Furthermore, as observed in Section 1.a, it has been chosen by EIOPA to change the definition of \bar{X} moving from a market mean (equal for all companies) to a company-specific mean. As a consequence, a factor entering into $\beta_{1,i}$ coefficient will correspond to the model regressor. Therefore, substituting the expressions for $\beta_0, \beta_{1,i}$ and $\beta_{2,i}$ into (2.23) one obtains the model⁴:

$$\widehat{\mathbf{Var}}_i(Y) = \beta^2 \sigma_i^2 \bar{X}_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, N. \quad (2.24)$$

⁴The result would not substantially differ if one chose as the independent variable a volume measure different from \bar{X}_i . For example, if X_{T_i} (the most recently observed value of X) would be chosen as the regressor, one would still have strong positive correlation between X_{T_i} and the factor \bar{X}_i entering into the $\beta_{1,i}$ coefficient. This however would suggest to redefine the model using a quadratic volume measure as regressor.

Denoting by $\bar{\sigma}^2 = \sum_{i=1}^N \sigma_i^2 / N$ the arithmetic mean of σ_i^2 on the whole sample of companies, model (2.24) can be approximated as:

$$\widehat{\mathbf{Var}}_i(Y) \approx \beta_2 \bar{X}_i^2 + \tilde{\varepsilon}_i, \quad i = 1, 2, \dots, N, \quad (2.25)$$

with $\beta_2 := \beta^2 \bar{\sigma}^2$. The approximation consists in assuming that the variability of σ_i^2 between companies (the parameter dispersion) can be well represented by considering it as included in the variance σ_ε^2 of the errors terms $\tilde{\varepsilon}_i$. So the problem just reduces to testing the assumption, through the linear regression (2.25), that the variance estimate function has a purely quadratic expression, i.e. with no constant and no linear terms.

For the variance estimate one could consider an approach similar to the “Standardised Method 1” proposed in QIS5 for the USPs for premium and reserve risk, using the estimator:

$$\widehat{\mathbf{Var}}_i(Y) = \bar{X}_i \frac{1}{T_i - 1} \sum_{t=1}^{T_i} X_{t,i} (Q_{t,i} - \bar{Q}_i)^2,$$

with:

$$Q_{t,i} := \frac{Y_{t,i}}{X_{t,i}} \quad \text{e} \quad \bar{Q}_i := \frac{\sum_{t=1}^{T_i} Y_{t,i}}{\sum_{t=1}^{T_i} X_{t,i}}.$$

However, the most consistent approach would be, where possible, to estimate the variance using the method of which we are currently testing the hypotheses⁵. This is equivalent to pose:

$$\widehat{\mathbf{Var}}_i(Y) = \hat{\sigma}_i^2(\hat{\delta}_i, \hat{\gamma}_i) \cdot \bar{Y}_i^2, \quad (2.26)$$

where $\hat{\sigma}_i^2$ is the estimate of the unit standard deviation of company i provided by Model M1, which shall be obtained by deriving the parameters $\hat{\delta}_i$ and $\hat{\gamma}_i$ after the minimisation of the corresponding criterion function.

Whatever is the variance estimator used, it is natural to apply the usual linear regression techniques to estimate – and validate – model (2.25), as discussed for testing the hypothesis on the mean. It cannot be excluded, however, that the parameter dispersion within the theoretical model can produce identification problems. It could be appropriate to exclude some outliers in the sample of the variance estimates.

2.a.3 An example of a market-wide test of M1V hypothesis

As an illustrative example, we performed a test on the variance hypothesis M1V for premium risk within the MTL segment based on the Italian

⁵This is the approach followed by the JWG to analyze the adequacy to market data of the different models used for the calibration (see [9], in particular par. 91, footnote 24).

market data⁶. The information used is publicly available on ANIA website www.infobila.it. We considered the time series from 1999 to 2013 of the earned premiums (variable X) and the corresponding ultimate cost estimate after the first development year (variable Y), observed on a selected sample of $N = 50$ companies operating within that market segment⁷. For each company i the parameters estimates $\hat{\delta}_i$ and $\hat{\gamma}_i$ have been computed that minimise the function $\sigma_i(\delta_i, \gamma_i)$, and the corresponding variance estimates (2.26) according to Model M1 have been obtained.

In a first run, the model (2.25) has been estimated on the sample of 50 companies, including an intercept. In the analysis of the results it is of primary importance the significance and the fitting ability of the model. One finds that the F statistic has a very high significance, which means that the model explains a significant portion of data variability; this outcome is confirmed by the high value of the R^2 and the R^2 adjusted by the degrees of freedom.

F -statistic	p -value	R^2	adj. R^2
737.89	< 0.0001	0.9389	0.9377

These results should be sufficient, by themselves, to support the acceptance of the M1V hypothesis on the used data. By performing the significance analysis also on the individual parameters we find that, consistently with the model assumptions, the intercept is not significantly different from zero and the coefficient of \bar{X}^2 is different from zero at a high significance level.

parameter	estimate	std. error	t -statistic	p -value
β_0	-78.11673	238.06326	-0.33	0.7442
β_2	0.00467	0.00017	27.16	< 0.0001

With a model-selection approach, we have estimated on the same 50 companies the reduced model with no intercept, obtaining the following results.

F -statistic	p -value	R^2	adj. R^2
841.59	< 0.0001	0.9450	0.9439

One can observe that the general significance and the fitting ability of the model further improve (both the F statistic and the R^2 are higher) and the

⁶It should be observed that a test of M1V hypothesis extended to the whole market is not necessarily more informative than a similar analysis made on a market segment. For example, if the market would be composed of two segments described by the same model but with different parameter values, the hypothesis testing would be more reliable if it was made only on the segment to which the considered company belongs.

⁷Companies with less than 5 observations and companies with $\hat{\sigma} \geq 1$ have been excluded.

parameter	estimate	std. error	<i>t</i> -statistic	<i>p</i> -value
β_2	0.00465	0.00016040	29.01	< 0.0001

slope coefficient β_2 is confirmed being different from zero with a high level of significance. We can conclude that the appropriateness of M1V hypothesis on the sample is largely and significantly confirmed.

2.a.4 Hypothesis on the distribution

We aim to test the M1D hypothesis, i.e. the assumption that the logarithms $\{\ln Y_t; t = 1, 2, \dots, T\}$ of the observations Y_t are a sample coming from a normal distribution. It should be emphasised that, given the small sample size which is typical in these applications, the normality tests can be problematic, since they can result of low significance or of low power (where “power” denotes the ability to avoid Type II errors, i.e. acceptance of normality when it is actually false). Problems in performing tests on the distributions have also been reported by the JWG in its calibration study⁸.

For the statistical testing of the M1D hypothesis both “algorithmic” and graphical methods can be considered.

Algorithmic Methods

Normality tests of algorithmic type assume the normality of data as the null hypothesis (H0), and define a test statistic which should allow to distinguish the null from the alternative hypothesis (H1), i.e. non-normality. In this approach, a low level of the *p*-value⁹ is, by definition, associated to a low level of confidence in normality of data. According to a common practice, *p*-value levels below 1% strongly support H1 (non-normality), levels above 10% indicate that data do not provide support to H1, while levels between 1% and 10% show an uncertainty condition. Therefore, as a preliminary remark, it should be emphasised that any normality test based on the *p*-value, no matter how large the data sample is, can possibly provide conclusive information for rejecting H0, but can also result to be inconclusive as concerning the acceptance of H0 (simply providing, in this case, no contrary evidence). This issue is well explained in [17]. The above mentioned difficulties faced by the JWG can be put in relation with this point, also.

Among the several logarithmic tests of normality presented in the literature, the following are the most commonly used.

⁸ *The empirical findings on this issue [i.e.: discriminating between the normal and log-normal distribution] – for example, with regard to the various goodness-of-fit diagnostics and PP-plots – were also inconclusive. [9], par. 102*

⁹Intuitively, the *p*-value is the probability of obtaining the data actually observed (and then obtaining for the test statistic the value actually computed or a more extreme value) when the null hypothesis H0 is actually true. Therefore a small *p*-value suggests to reject H0, but a large *p*-value does not exclude that the alternative hypothesis H1 is also true.

- **Kolmogorov-Smirnov test.** This is a non-parametric test based on the Empirical Distribution Function (EDF). Given a sample $\{X_1, X_2, \dots, X_n\}$ of n independent and identically distributed (i.i.d.) observations of the random variable X , the EDF of X is defined as:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{X_i \leq x\}}, \quad x \in \mathbb{R}.$$

Given a theoretical continuous distribution function $F(x)$ which is assumed as the true distribution (in this case the normal distribution), the goodness-of-fit of the sample with respect to $F(x)$ is defined introducing a *distance measure* between the empirical distribution function $F_n(x)$ and the theoretical distribution function. In the Kolmogorov-Smirnov test (KS) [22], the distance measure is defined as the supremum D of the difference, in absolute value, between $F_n(x)$ and $F(x)$:

$$D = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Obviously, the lower the value of D the stronger the support provided to H_0 hypothesis. Numerically, the test consists in comparing $\sqrt{n} D$ with the corresponding Kolmogorov critical value K_α , with K_α such that $\mathbf{P}(K \leq K_\alpha) = 1 - \alpha$, where K is the Kolmogorov random variable and α is the chosen significance level. As for all tests based on a distance measure, the p -value is the probability that D is *greater* than the observed value. In practice, the KS statistic requires a relatively large number of observation in order that the null hypothesis is properly rejected.

- **Cramer-von Mises test.** The Cramer-von Mises test (CvM) is also based on the EDF, however it belongs to the class of Quadratic EDF (QEDF) statistics. These tests use a quadratic distance measure, defined as:

$$D^2 = n \omega^2, \quad \text{with} \quad \omega^2 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 w(x) dF(x),$$

where $w(x)$ is a fixed weight function. Compared to the KS tests, the QEDF type tests take better into account the whole data in the sense of the sum of the variations, while the KS test is more sensitive to the aberrance in the sample.

The CvM tests [3] [30] uses D^2 with $w(x) \equiv 1$:

$$T^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x).$$

It consists in comparing T^2 with the corresponding tabulated value, at a given level of significance α . In normality tests, CvM should display high

power, being one of the most efficient EDF tests in detecting departures from the null hypothesis (low rate of Type II errors). The use of this test is usually recommended for samples with $n < 25$ (while it can fail with very large samples).

- **Anderson-Darling test.** The Anderson Darling test (AD) [1] is also in the QEFD class, with the weight function given by:

$$w(x) = 1/[F(x)(1 - F(x))],$$

then the test statistic is:

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x).$$

The properties are similar to CvM, with the only difference that the A^2 statistic gives more weight to the tails.

- **Shapiro-Wilk test.** The Shapiro-Wilk normality test (SW) [26] compares a variance estimator based on the optimal linear combination of the order statistics of a normal variable and the usual sample variance. The test statistic W is the ratio between these two estimators and its value can range between 0 and 1. The normality hypothesis is rejected for low values of W and not rejected for values close to 1. Therefore, the p -value is the probability that W is lower than the observed value. It should be pointed out, however, that the distribution of W is highly asymmetric, so much that W values close to 0.9 can be considered to be low in the normality analysis.

For interpreting the results, it can be useful to observe that the W statistic can be interpreted as the square of the correlation coefficient in a QQ-plot. The SW test is often presented as one of the most powerful test for normality in small samples. It could be unreliable if there are many repeated values in the data (tied observations).

- **Jarque-Bera test.** This test belongs to the *omnibus moments* class, as it assesses simultaneously whether two sample moments, the skewness and kurtosis, are consistent with the normality assumption.

The Jarque-Bera test statistic (JB) [16] has the following expression:

$$JB = \frac{T}{24} (4b + (k - 3)^2),$$

where \sqrt{b} and k are, respectively, the sample skewness and kurtosis. For normal data the JB statistic asymptotically has a chi-squared distribution with two degrees of freedom.

In the JB test, H_0 is a joint hypothesis of both the skewness and the excess kurtosis being zero. This hypothesis is rejected for high JB values.

Therefore the p -value is the probability of JB being higher than the observed value.

The JB test has been used by the JWG to identify outliers within the standard deviation estimates obtained on a relatively large sample of companies. The test, however, is not appropriate for small samples, since the chi-squared approximation is overly sensitive¹⁰ and, moreover, the distribution of p -values becomes a right-skewed uni-modal distribution. These behaviours tend to produce a high level of Type I errors (the null hypothesis is improperly rejected). For all the above mentioned reasons, it doesn't seem appropriate to use the JB test for the problem we are considering here.

Graphical Methods

- **Histogram.** This is the usual bar-chart that illustrates the relative frequency of observations falling into the k -th interval of a “grid” properly defined on the x axis. Given the limited number of observations available in USP computations, generally this approach is of little practical use in checking for normality of data.
- **PP-plot.** Given a sample $\{X_1, X_2, \dots, X_n\}$ of n independent and equally distributed observations of the random variable X , let us derive the ordered sample $\{X_{n,n} \leq X_{n-1,n} \leq \dots \leq X_{1,n}\}$. Since $X_{k,n} \leq x$ if and only if $\sum_{i=1}^n \mathbf{I}_{\{X_i > x\}} < k$, on the ordered sample the EDF takes on the values:

$$F_n(X_{k,n}) = \frac{n - k + 1}{n}, \quad k = 1, 2, \dots, n.$$

The *probability plot* (PP-Plot) is the two-dimensional graph:

$$\left\{ \left(F(X_{k,n}), \frac{n - k + 1}{n + 1} \right), k = 1, 2, \dots, n \right\},$$

built on the ordered sample $\{X_{n,n} \leq X_{n-1,n} \leq \dots \leq X_{1,n}\}$ of the n (i.i.d.) observations of X .

By Glivenko-Cantelli theorem, if X has distribution function F the plot should be approximately linear.

- **QQ-plot.** The *quantile plot* (QQ-plot) is the same graph referred to quantiles:

$$\left\{ \left(X_{k,n}, F^{(-1)} \left(\frac{n - k + b_k}{n + a_k} \right) \right), k = 1, 2, \dots, n \right\},$$

¹⁰The problem is mentioned also by the JWG. In [9], par. 9.3 it is said: *Care should be exercised with this test statistic as the asymptotic distribution only holds for fairly large ($n \gg 100$) numbers of observations n .*

with a_k and b_k appropriately chosen to take into account the empiric discontinuity of the distribution (see e.g. [10]). Typical choices are $a_k \equiv b_k \equiv 1$, or $a_k \equiv 0$ and $b_k \equiv 0.5$. Also in this case, if $X \sim F$ the plot should be approximately linear.

2.a.5 Comparing the testing methods

We performed a comparative analysis through a simulation exercise with the aim to compare the discriminant ability of the normality tests previously considered, with a specific attention to small samples.

Organization of the simulation exercise

The tests KS, CvM, AD, SW and JB have been applied to 1000 samples of T observations (with $T = 6, 10, 15, 100$) drawn by simulation from:

- a normal distribution,
- a lognormal distribution,
- a Weibull distribution with shape parameter $\tau > 1$, and
- a Pareto Type II distribution.

For all distributions we set a mean $m = 100$ and for the normal distribution we chose a variation coefficient $\kappa = 0.1$ (which is a typical figure for the unit standard deviations prescribed in the standard formula). It follows that the quantile of the normal at probability level $p = 99.5\%$ is $Q_p = 125.758$. The parameters of the other three distributions have been chosen in order to have the same value for Q_p (therefore, the same value for the unexpected loss)¹¹. As a result, the Weibull distribution has shape parameter $\tau = 9.4315$ and scale parameter $\theta = 105.3799$, that imply a standard deviation $\sigma = 12.71$; then the dispersion is higher than in the normal distribution (where $\sigma = 10$), in line with the fact that for $\tau > 1$ the Weibull distribution is more light-tailed. The lognormal and the Pareto distribution have a lower dispersion compared to the normal, since both the distributions, in particular Pareto,

¹¹The Weibull distribution function has the form:

$$F(x) = 1 - e^{-(x/\theta)^\tau}, \quad x > 0,$$

with $\theta, \tau > 0$. The mean and the p -quantile are:

$$\mu = \theta \Gamma(1 + 1/\tau), \quad Q_p = \theta [-\ln(1 - p)]^{1/\tau}.$$

For the Pareto Type II distribution (also referred to as Lomax distribution) one has:

$$F(x) = 1 - \left(\frac{\theta}{\theta + x} \right)^\alpha, \quad x > 0,$$

with $\alpha, \theta > 0$. The mean and the p -quantile are given by:

$$\mu = \frac{\theta}{\alpha - 1}, \quad Q_p = \theta [(1 - p)^{-1/\alpha} - 1].$$

For the properties of the Weibull and Pareto Type II distribution see e.g. [17].

are more heavy-tailed. In particular, for the lognormal the mean and the standard deviation (of $\ln Y$) are $\mu = 4.60$ and $\omega = 0.09058$, which implies a standard deviation $\sigma = 9.07$. For the Pareto distribution, where the kurtosis is much higher, one has a shape parameter $\alpha = 1.0065$ and a scale parameter $\theta = 0.6542$ (for these values the variance does not exist).

In summary, besides data generated from a normal distribution (which corresponds to the null hypothesis), we have considered three alternative hypotheses, one corresponding to a lower kurtosis (Weibull), and two (lognormal and Pareto) corresponding to a higher kurtosis, one of which (Pareto) has extreme behaviour. We imposed to all the distributions the same value of the unexpected loss in order to make the four alternatives equivalent from the point of view of the implied SCR, as defined in Solvency II.

Simulations results

Algorithmic methods. The 1000 values of each test statistic and the corresponding p -values, computed in each simulation on the samples with $T = 6, 10, 15, 100$ observations, have been saved and compared within each other. An exhaustive analysis of the results can be obtained by a systematic comparison of the empirical distributions thus derived. We report here the results of a reduced analysis, which only takes into account the mean, the mode and the median of the distributions as well as the number of rejections of the null hypothesis.

In Tables 1a and 1b we reported, for all the sample sizes considered, the simulation results of the five normality tests previously illustrated. Table 1a refers to the Kolmogorov-Smirnov, the Cramer-von Mises and the Anderson-Darling test, which are based on distance measures; Table 1b concerns the Shapiro-Wilk and the Jarque-Bera test. In both tables the following figures are reported: the mean of the test statistic, the mean of p -value, the mode¹² of p -value, the median of p -value and the rejection rate of H_0 at level α , that is the percentage number r_α of cases, out of the 1000 simulated cases, where the p -value was lower than the significance level α ; the levels $\alpha = 1\%, 5\%, 10\%$ have been considered. In the normal case the rejection rate r_α provides the rate of Type I errors (H_0 is rejected when it is in fact true); obviously one requires that the value of r_α is as low as possible. In the three cases of non-normality, instead, r_α should be as high as possible, since it provides a measure of the power of the test (i.e. the ability of rejecting the H_0 hypothesis when it is false); it is generally accepted that on non-normal data r_α should be 80% or greater. The complement to unity of r_α for the non-normal data provides the Type II error rate (failure to reject H_0 when it is false).

The figures reported in the tables show that the Type I error rate is appropriately low for all the five tests and for all the values of sample size

¹²The mode has been computed by rounding to the third decimal place the simulated p -value. In the case of multiple values the minimum value has been taken.

			Shapiro-Wilk						Jarque-Bera						
	stat	p-value			tasso di rifiuto			stat test	p-value			tasso di rifiuto			
distr	T	media	moda	mediana	$\alpha=1\%$	$\alpha=5\%$	$\alpha=10\%$		media	moda	mediana	$\alpha=1\%$	$\alpha=5\%$	$\alpha=10\%$	
Nor	6	0.906	0.491	0.447	0.489	1.0%	4.4%	9.1%	1.460	0.597	0.658	0.645	1.2%	3.5%	5.7%
Nor	10	0.926	0.494	0.651	0.495	1.0%	4.0%	9.2%	1.488	0.616	0.000	0.661	1.5%	3.9%	6.2%
Nor	15	0.944	0.527	0.170	0.536	1.0%	3.3%	7.9%	1.527	0.622	0.000	0.663	1.7%	4.1%	5.6%
Nor	100	0.987	0.504	0.478	0.497	1.4%	4.0%	8.5%	1.902	0.543	0.000	0.570	2.0%	4.0%	6.3%
Log	6	0.905	0.486	0.541	0.482	1.2%	4.5%	9.5%	1.481	0.594	0.647	0.631	1.3%	4.2%	6.3%
Log	10	0.924	0.484	0.062	0.475	1.1%	5.4%	10.5%	1.583	0.605	0.000	0.650	2.1%	3.9%	7.2%
Log	15	0.942	0.510	0.369	0.508	1.1%	4.6%	10.1%	1.761	0.605	0.000	0.655	3.0%	5.5%	7.9%
Log	100	0.984	0.402	0.008	0.351	4.7%	14.4%	21.9%	3.324	0.435	0.000	0.421	7.4%	13.4%	17.9%
Wei	6	0.903	0.490	0.135	0.488	1.9%	6.3%	12.4%	1.671	0.584	0.019	0.636	2.1%	6.2%	9.4%
Wei	10	0.917	0.443	0.277	0.419	2.1%	7.8%	14.9%	2.006	0.572	0.000	0.622	4.5%	7.3%	10.0%
Wei	15	0.930	0.434	0.004	0.397	3.5%	11.9%	19.4%	2.931	0.533	0.000	0.580	7.5%	11.6%	13.9%
Wei	100	0.967	0.110	0.000	0.021	39.8%	62.7%	73.3%	11.223	0.131	0.000	0.044	35.8%	51.8%	62.4%
Par	6	0.748	0.128	0.000	0.024	40.1%	57.3%	67.0%	6.595	0.243	0.001	0.088	36.8%	46.6%	51.1%
Par	10	0.658	0.031	0.000	0.000	74.2%	87.6%	91.7%	24.361	0.104	0.000	0.000	60.9%	68.8%	73.9%
Par	15	0.589	0.007	0.000	0.000	91.5%	96.6%	98.2%	72.505	0.032	0.000	0.000	83.0%	87.9%	90.6%
Par	100	0.304	0.000	0.000	0.000	100.0%	100.0%	100.0%	15733.114	0.000	0.000	0.000	100.0%	100.0%	100.0%

Table 1b. Results of Shapiro-Wilk and Jarque-Bera tests on simulated data

T . In detail, the value of r_α for normal data is higher in the SW and JB tests than in the three tests based on distance measures, and among the latter KS, in turn, seems to provide lower values.

If one looks, however, at the H_0 rejection rates for the non-normal data, for all the tests one observes inappropriately low r_α values both on the log-normal and the Weibull data, for all three levels of α . Sufficiently high levels of the rejection rate can be found only for data with Pareto distribution, but also in this case r_α values greater than 80% can be observed only for high values of T and α .

All the test statistics have the theoretically expected behavior: as data departs from normality, one observes a decreasing trend for the SW test (consistently with the interpretation of the W statistic as the squared correlation coefficient in the QQ-plot) and an increasing trend for the other tests. However, also for the sample with $T = 100$ all the tests almost systematically fails to detect non-normality for distributions which are not heavily different (as the Pareto) from a bell-shaped distribution. By and large, also taking into account mean, mode and median of the p -values, one can conclude perhaps that SW and AD tests are slightly more powerful; there is, however, a high probability of Type II errors for all the methods considered. Among the five cases considered the JB test seems the worst performing, probably because of the small size of the samples considered. The indication is then confirmed of not using this method for this kind of applications.

è [*Graphical methods*. In order to compare the performances also of normality tests of graphic type, for each value of T the sample has been selected, out of the 1000 simulated, where the p -value for a given test statistic is closest to the value of the mode. The Shapiro-Wilk statistic has been used. For each of these samples a PP-plot and a QQ-plot has been produced; these plots are reported in Tables 2 and 3. With the exception of the extreme case of Pareto data, all these graphical tests confirm the difficulty in correctly identifying the normal data for small samples. With this data it is hard to discriminate using a PP-plot between the normal and the lognormal hypothesis also on the sample with 100 observations.

2.a.6 Appropriateness of maximum likelihood method

As concerning the ML property, i.e. the appropriateness of the maximum likelihood method used for the estimate (Section 2.a), this is testified by the convergence properties of the minimisation procedure, which is required to univocally identify a minimum of the criterion function in the optimisation interval D . The uniqueness of the minimum provided by the procedure can be tested by an empirical illustration of the criterion function $\ell(\delta, \gamma)$ on a sufficiently large grid of δ and γ values.

In order to define the grid the values γ_{min} and γ_{max} have to be chosen. Recalling the definition $\gamma = \ln(\sigma/\beta)$ and since $\beta = \mathbf{E}(Y/X)$ (expected loss

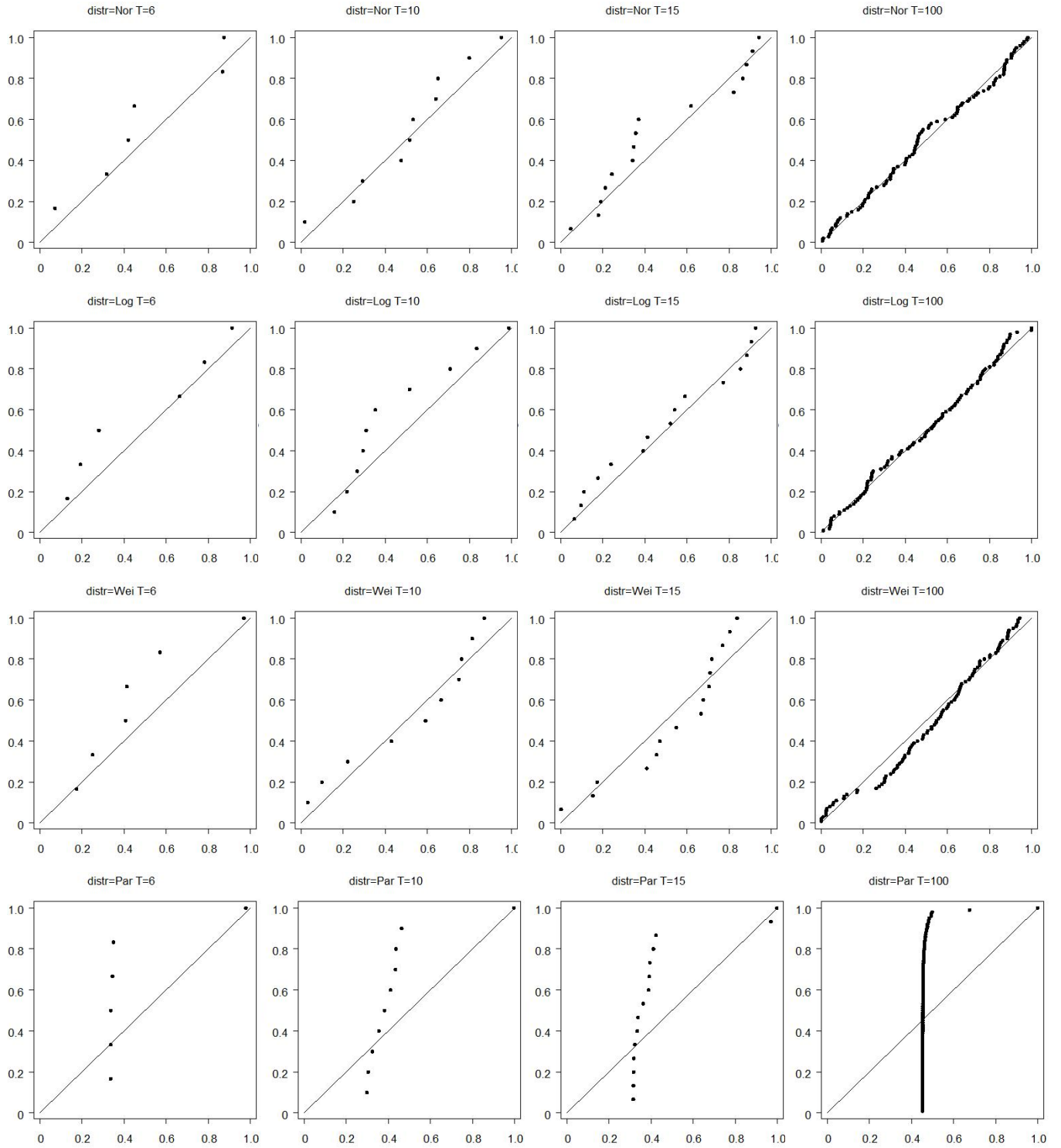


Table 2. PP-plots on samples with modal p -value (according to Shapiro-Wilk)

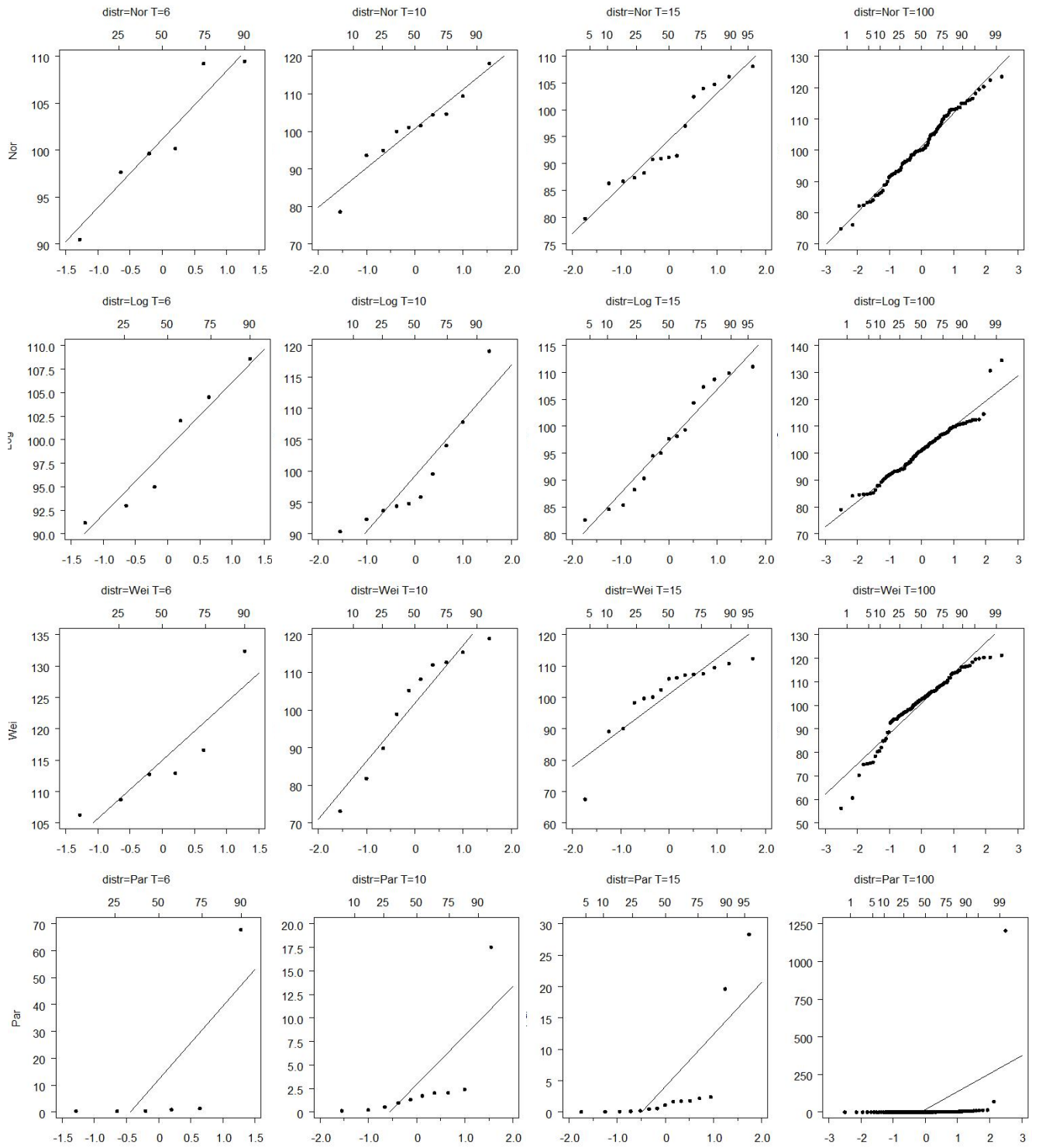


Table 3. QQ-plots on samples with modal p -value (according to Shapiro-Wilk)

ratio, expected run-off ratio) one can assume $\beta \approx 1$ and $\sigma \in [0.005, 1]$, hence:

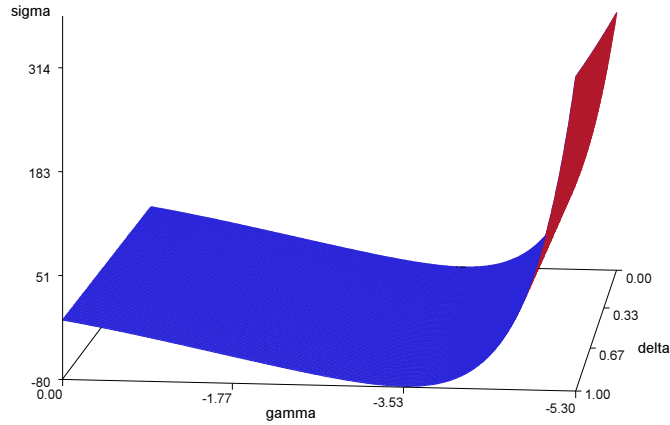
$$\gamma_{min} = \ln(0.005) = -5.30, \quad \gamma_{max} = \ln(1) = 0.$$

So the domain D is restricted to the domain:

$$D^* = \{0 \leq \delta \leq 1, \gamma_{min} \leq \gamma \leq \gamma_{max}\},$$

with the assumption that values of the criterion function outside this interval are irrelevant for the analysis. A three-dimensional graph on D^* should show the regularity of the function and the existence of a global minimum (possibly on the boundary of D^*) clearly identified by the minimisation procedure. As an example, a typical “volatility surface” $\sigma(\delta, \gamma)$ is illustrated in Figure 1.

Figure 1. The surface $\sigma(\delta, \gamma)$ on the domain D^*



2.b Hypothesis testing for Model M2

Also for Model M2 it is required by the Delegated Acts to verify the consistency between the underlying assumptions and the data. Specifically, a statistical testing is required for the assumptions introduced in Section 1.b:

M2I – *Independence hypothesis*: independence between cumulated (and incremental) payments of different accident years (AY);

M2M – *Hypothesis on the conditional mean*: for any AY and in any development year (DY) of any given AY, proportionality of the expected cumulative payments of next DY with respect to the cumulative payments of current DY;

M2V – *Hypothesis on the conditional variance*: for any AY and in any DY of any given AY, proportionality of the variance of cumulative payments of next DY with respect to the cumulative payments of current DY.

It is useful to rewrite hypotheses M2M and M2V in a unified form. Let us denote by:

$$\mathcal{B}_0 := \{C_{0,0}, C_{0,1}, \dots, C_{0,I}, \},$$

the set of all payments made in the first development year. Then the assumptions M2M and M2V can be unified as:

M2MV - *Time series hypothesis*. There exists constants $f_j > 0$ and $\sigma_j > 0$ and random variables $\varepsilon_{i,j}$ such that for $1 \leq j \leq J$ and $0 \leq i \leq I$:

$$C_{i,j} = f_{j-1} C_{i,j-1} + \sigma_{j-1} \sqrt{C_{i,j-1}} \varepsilon_{i,j}, \quad (2.27)$$

where $\varepsilon_{i,j}$ are error terms identically distributed and conditionally independent, given \mathcal{B}_0 , with mean $\mathbf{E}(\varepsilon_{i,j}|\mathcal{B}_0) = 0$ and variance $\mathbf{Var}(\varepsilon_{i,j}|\mathcal{B}_0) = 1$.

This formulation has been proposed in 2006 by Buchwalder, Bühlmann, Merz and Wüthrich [4] as a distributional extension of Mack's DFCL model and defines the so-called Time Series Chain Ladder (TSCL) model. Expression (2.27) allows, among other things, a simulative approach to the model¹³.

2.b.1 Hypothesis on the conditional mean and variance

For any fixed $j = 0, 1, \dots, J-1$, expression (2.27) defines a linear regression model for the observations of a pair of consecutive development years. Precisely, one has J *weighted* linear regressions of the type:

$$y_i = \beta x_i + \frac{\sigma}{\sqrt{w_i}} \varepsilon_i, \quad i = 0, 1, \dots, I,$$

with $x_i = C_{i,j-1}$, $y_i = C_{i,j}$ and $w_i = 1/x_i = 1/C_{i,j-1}$. As it is well-known, the β coefficient in this regression will be estimated by weighted least squares as:

$$\hat{\beta} = \frac{\sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i x_i^2};$$

as one immediately checks, this expression coincides with (1.16), which provides the chain ladder estimator \hat{f}_j . Moreover, the variance of the error terms is estimated as:

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n-1},$$

¹³By a strictly theoretical point of view, the recursive relation defined by the time series assumption could produce negative values for the cumulative payments $C_{i,j-1}$. This “negativity problem”, just extensively discussed in the *Comments* to the original paper, could be avoided by reformulating the properties of the error terms $\varepsilon_{i,j}$ conditionally on the value taken by $C_{i,j-1}$. This would lead to a model with a much more complex dependence structure. Since the negativity problem is usually irrelevant in practical applications, in the TSCL one takes the pragmatic position of ignoring this theoretical inconsistency.

where the SSE has the form:

$$\text{SSE} := \sum_{i=1}^n w_i (y_i - x_i \hat{\beta})^2;$$

this expression coincides in turn with the estimator $\hat{\sigma}_j^2$ of DFCL given by (1.17).

The important point here is that, in order to assess the significance of these estimates and the consistency of data to the model, one can use the traditional tests for the hypotheses and for the goodness-of-fit, thus obtaining a test for M2MV, i.e. both for the hypothesis M2M on the conditional mean and for the hypothesis M2V on the conditional variance. Therefore one will just have to perform (with the proper changes to account for homoscedasticity) the F test and/or the t test, with the corresponding p -value, and compute measures of fit (SSE) and of explained variance (R^2). If a preliminary analysis is performed considering a model with intercept, this should result significantly not different from zero. A graphical illustration could be added.

Remark. All these measures of significance and goodness-of-fit are included in the test plan provided by the procedure Explorer[©], which is aimed at performing an exploratory analysis of data in connection with a variety of loss reserving models. One of the most early works on the goodness-of-fit methods applied to loss reserving models has been proposed in 1998 by Venter [27]; for further developments see e.g. [7]. ■

Usually, data will not be sufficient to perform all the J regressions which are theoretically required. In fact the number of observations (accident years) available for estimating regression j is $I - j$, therefore decreases when j increases. For example, if one decides that at least 5 observations are needed in order that a regression is considered significant, the analysis for the last development year will be performed only if a trapezoid of paid losses with $I \geq J + 4$ is available. In the usual case of a triangle ($I = J$), only the first $I + 1 - 5$ regressions could be considered.

2.b.2 Independence hypothesis. Test on time series residuals

One of the methods available for testing independence between different accident years consists in testing the independence of the residuals derived by the time series equation (2.27). The basic idea, also proposed by Merz and Wüthrich [31], is to verify through a linear regression that there are not trends over accident years in these residuals.

Considering the individual development factors $F_{i,j}$, expression (2.27) can be written:

$$F_{i,j} := \frac{C_{i,j}}{C_{i,j-1}} = f_{j-1} + \frac{\sigma_{j-1}}{\sqrt{C_{i,j-1}}} \varepsilon_{i,j}, \quad (2.28)$$

where $\varepsilon_{i,j}$ are, by assumption, identically distributed and conditionally independent, given \mathcal{B}_0 , with zero conditional mean zero and unit conditional variance. Hence, if also the assumption M2I of independence between accident years holds, the random variables:

$$\epsilon_{i,j} := \frac{F_{i,j-1} - f_{j-1}}{\sqrt{\sigma_{j-1}^2 C_{i,j-1}^{-1}}},$$

are independent. Then, in order to verify hypothesis M2I one can test the independence of the $\epsilon_{i,j}$ on the observed trapezoid, that is the independence of the *residuals*:

$$\begin{aligned} \epsilon_{i,j} &= \frac{F_{i,j-1} - f_{j-1}}{\sqrt{\sigma_{j-1}^2 C_{i,j-1}^{-1}}}, \\ j &= 1, 2, \dots, J, \quad i = 0, 1, \dots, I - j. \end{aligned} \quad (2.29)$$

The number of these residuals is $n^{TS} = J(I - J) + J(J + 1)/2$.

However the residuals given by (2.29) are not observable, since the parameters f_j e σ_j are not known. Replacing in expression (2.29) the unknown parameters by the parameter estimates obtained by (1.16) and (1.17) one then obtains the n^{TS} observable residuals:

$$\begin{aligned} \hat{\epsilon}_{i,j}^{TS} &= \frac{F_{i,j-1} - \hat{f}_{j-1}}{\sqrt{\hat{\sigma}_{j-1}^2 C_{i,j-1}^{-1}}}, \\ j &= 1, 2, \dots, J, \quad i = 0, 1, \dots, I - j, \end{aligned} \quad (2.30)$$

on which an independence test can be actually performed.

The independence between residuals of different accident years for the same development year has just been implicitly tested in the regression analysis for testing the M2MV hypothesis. It is required more here, since we need to explicitly test the independence between residuals of different accident years and of *any* development year. This independence test can be made through a graphical analysis. If the independence assumption holds, we should not observe any trend over the accident years in the residual plot. The absence of trends can also be checked by a regression analysis, performed by development year¹⁴ or, more simply, on the whole sample of residuals.

Remark. Expression (2.30) can also take the form:

$$\begin{aligned} \hat{\epsilon}_{i,j}^{TS} &= \frac{C_{i,j} - \hat{C}_{i,j-1}^{TS}}{\hat{\sigma}_{j-1} \sqrt{C_{i,j-1}}}, \\ j &= 1, 2, \dots, J, \quad i = 0, 1, \dots, I - j, \end{aligned} \quad (2.31)$$

¹⁴In this case the remarks still hold on the minimum number of observations required for testing the M2MV hypothesis: the residual analysis shall be performed only for the development years having a sufficient number of observations.

where:

$$\widehat{C}_{i,j-1}^{TS} := \hat{f}_{j-1} C_{i,j-1},$$

can be interpreted as the *fitted value* of the TSCL model. ■

Problems of spurious dependence

It is important to observe that the results of the independence tests on the residuals could be distorted by spurious dependence effects, induced by the use of the chain ladder estimators \hat{f}_j . In particular, if one considers the “column” linear combinations $\sum_{i=0}^{I-j} \sqrt{C_{i,j-1}} \widehat{\epsilon}_{i,j}^{TS}$, one finds that the following relations hold:

$$\sum_{i=0}^{I-j} \sqrt{C_{i,j-1}} \widehat{\epsilon}_{i,j}^{TS} = 0, \quad j = 1, 2, \dots, J. \quad (2.32)$$

In fact, using expression (2.31) for the time series residuals, one has, for $j = 1, 2, \dots, J$:

$$\begin{aligned} \sum_{i=0}^{I-j} \sqrt{C_{i,j-1}} \widehat{\epsilon}_{i,j}^{TS} &= \sum_{i=0}^{I-j} \sqrt{C_{i,j-1}} \left(\frac{C_{i,j} - \hat{f}_{j-1} C_{i,j-1}}{\hat{\sigma}_{j-1} \sqrt{C_{i,j-1}}} \right) \\ &= \frac{1}{\hat{\sigma}_{j-1}} \sum_{i=0}^{I-j} (C_{i,j} - \hat{f}_{j-1} C_{i,j-1}) \\ &= \frac{1}{\hat{\sigma}_{j-1}} \left(\sum_{i=0}^{I-j} C_{i,j} - \hat{f}_{j-1} \sum_{i=0}^{I-j} C_{i,j-1} \right) = 0, \end{aligned}$$

where the last equality follows from (1.16).

Expressions (2.32) show that there exist negative correlations between the observable time series residuals (i.e. calculated with the estimated parameters). In particular, for $j = I - 1$ one finds that $\widehat{\epsilon}_{0,I-1}^{TS}$ and $\widehat{\epsilon}_{1,I-1}^{TS}$ are negatively perfectly correlated.

An additional issue, however less important, is that expressions (2.32) imply the property:

$$\sum_{j=1}^J \sum_{i=0}^{I-j} \sqrt{C_{i,j-1}} \widehat{\epsilon}_{i,j}^{TS} = 0, \quad (2.33)$$

which is incompatible with the property:

$$\frac{1}{n^{TS}} \sum_{j=1}^J \sum_{i=0}^{I-j} \widehat{\epsilon}_{i,j}^{TS} = 0. \quad (2.34)$$

Then the distribution of the observed residuals has not zero mean.

Remark. Properties (2.32) have just been derived by Merz and Wüthrich in [31], Section 7.4, together with the variance properties:

$$\mathbf{Var}(\hat{\epsilon}_{i,j}^{TS} | \mathcal{B}_{j-1}) = 1 - \frac{C_{i,j-1}}{\sum_{i=0}^{I-j} C_{i,j-1}} < 1, \quad (2.35)$$

(with $\mathcal{B}_k := \{C_{i,j}; i+j \leq I, 0 \leq j \leq k\}$), which imply that the variance of the empirical distribution of residuals is lower than the unit theoretical value. All these properties of the empirical residuals have been used by the authors in connection with the simulation of the TSCL model by parametric bootstrap. ■

2.b.3 Independence hypothesis. Test on Pearson residuals

An alternative to using time series residuals is to consider (unadjusted) *Pearson residuals*¹⁵:

$$\epsilon_{i,j}^P = \frac{X_{i,j} - X_{i,j}^{fit}}{\sqrt{X_{i,j}^{fit}}}, \quad (2.36)$$

$$j = 0, 1, \dots, J, \quad i = 0, 1, \dots, I-j,$$

where $X_{i,j}^{fit}$ are the *fitted* incremental payments obtained by *backcasting* from the last observed diagonal. Precisely, one defines the *fitted* cumulative payments by the backward recursive procedure:

$$C_{i,j}^{fit} = \frac{C_{i,I-j}}{f_j f_{j+1} \cdots f_{I-j-1}}, \quad (2.37)$$

and then derives as usual, by differencing, the corresponding incremental payments $X_{i,j}^{fit}$. Pearson residuals $\epsilon_{i,j}^P$ are widely used in the *generalized linear model* theory (GLM) and for this reason they are usually chosen as “noise generators” in the bootstrap simulation of the stochastic chain ladder model, when this is specified in the form of an Over Dispersed Poisson model (ODP). It can be immediately checked, in fact, that the ODP model can be reformulated as a GLM model (see e.g. [12], [11], [5]). In this theoretical framework, the Pearson residuals $\epsilon_{i,j}^P$ have zero mean and constant variance (equal to the ODP overdispersion parameter ϕ).

As pointed out by Verral and England in [28] and [29], the backward fitted values given by (2.37) are the most appropriate for defining the residuals of a recursive model like the chain ladder, and they have better theoretical

¹⁵The adjusted (i.e. corrected for the number of degrees of freedom) Pearson residuals are obtained by multiplying the unadjusted residuals by $\sqrt{n^{TS}/(n^{TS} - p)}$, where p is the number of parameters. This adjustment is irrelevant for the purpose of testing independence.

properties then the fitted values of the type $C_{i,j}^{TS} = f_{j-1}C_{i,j-1}$ used for the time series residuals.

The number of the Pearson residuals is $n^P = J(I - J) + (J + 1)(J + 2)/2$, which is greater than the number of the time series residuals. Moreover, while the time series residuals are adimensional variables (they are pure numbers), Pearson residuals have dimension $\text{euro}^{1/2}$ (the squared residuals have monetary dimension) and then take on numerical values on a different scale.

Obviously, also the residuals (2.36) are not observable and their observable version is obtained by the estimates:

$$\hat{\epsilon}_{i,j}^P = \frac{X_{i,j} - \hat{X}_{i,j}^{fit}}{\sqrt{\hat{X}_{i,j}^{fit}}}, \quad (2.38)$$

$$j = 0, 1, \dots, J, \quad i = 0, 1, \dots, I - j,$$

where the estimate of the fitted incremental payments $\hat{X}_{i,j}^{fit}$ is obtained by (2.37) replacing the unknown development factors f_j by the corresponding chain ladder estimators \hat{f}_j . Therefore one can suppose that also the observed Pearson residuals $\hat{\epsilon}_{i,j}^P$ contain spurious correlation induced by the use of these estimators, even though one could argue that the use of a product instead of a single estimator should induce more weak correlations. For the Pearson residuals, however, the theoretical analysis of these effects is more difficult than for the time series residuals and the performances of the two methods could be better compared using empirical approaches. A useful comparison can be obtained by simulation, generating a sample of “pseudotrapezoids” of independent payments and analyzing the two types of residuals estimated on each pseudotrapezoid.

Part II

Application to entity-specific data

3 Premium Risk – Model M1

3.a Specification of the input data

In the *Premium Risk* submodule the undertaking-specific unit standard deviation for each *segment* (a specified group of lines of business, as defined in [14]) can be estimated only using Model M1, which is summarised in Section 1.a. In this application of the model the data used consists of:

- Y_t : the *aggregated losses* of accounting year t , with $t = 1, \dots, T$ and $T \geq 5$, that is the sum of the payments and the best estimate provisions made at the end of year t for claims occurred and reported in the same year;
- X_t : the earned premiums of accounting year t , with $t = 1, \dots, T$ and $T \geq 5$.

In general, available data concerns different types of insurance activity (direct business, accepted business, direct plus accepted business), net or gross of recoveries from policyholders (deductibles, salvages, subrogations). In this application it seems appropriate to use data concerning direct plus accepted business net of recoveries.

As concerning outward reinsurance, data used will be net or gross of reinsurance recoverables according to whether the market-wide adjustment factor NP_{MW} for non-proportional reinsurance([13], art. 117(3))¹⁶ is used or not.

It is required that data are representative for the premium risk which the company will face in the twelve month following the valuation date (the “next year”, i.e. year $t = T + 1$).

The claims cost Y_t is given by:

$$Y_t = P_t + R_t - (P_t^r + R_t^r) - \Delta_t^r,$$

where (in brackets it is reported the entry of the IVASS supervisory form – *modulo di vigilanza* – n.17 relative to, e.g., the gross direct business):

¹⁶As previously pointed out, we do not consider in this paper the entity-specific adjustment factor NP_{USP} ([13], art. 218(1.iii)).

- P_t : paid amounts for claims occurred in accounting year t (v10);
- R_t : claims provision at the end of accounting year t for claims occurred in the same year (v13);
- P_t^r : amounts recovered in accounting year t for deductibles, salvages and subrogations from policyholders and third parties for claims occurred in the same year (v14);
- R_t^r : amounts to be recovered for deductibles, salvages and subrogations from policyholders and third parties at the end of accounting year t for claims occurred in the same year (v15);
- Δ_t^r : balance of portfolio movements for claims occurred in the same year (v17).

Remark. For the MTL segment the amounts P_t includes the contributions to F.G.V.S. fund (v301). ■

The claims cost can be adjusted by excluding catastrophe claims cost to the extent that the risk of such claims is just considered in the subforms specific for that.

The earned premiums X_t are defined by:

$$X_t = EP_t = R_{t-1}^p + WP_t - R_t^p + \Delta_t^p + \Delta_t^{cp},$$

where:

- R_{t-1}^p : premium provision at the end of previous accounting year $t - 1$ (v01);
- WP_t : written premiums in accounting year t (v03);
- R_t^p : premium provision at the end of accounting year t (v04);
- Δ_t^p : balance of portfolio movements for premiums received in accounting year t (v05);
- Δ_t^{cp} : balance of net exchange differences deriving from the updating of foreign currency provisions in accounting year t (v02).

3.b Application of the method

The undertaking-specific unit standard deviation for segment s according to Method 1 is given by:

$$\sigma_{(prem,s,USP)} = c \cdot \hat{\sigma}(\hat{\delta}, \hat{\gamma}) \cdot \sqrt{\frac{T+1}{T-1}} + (1-c) \cdot \sigma_{(prem,s)},$$

where:

- T is the length in years of the yearly time series;
- c is the credibility factor;

- $\sigma_{(prem,s)}$ is the market-wide level of the unit standard deviation, net of reinsurance, prescribed by EIOPA; this coefficient is obtained multiplying by NP_{MW} the gross standard deviation;
- $\hat{\sigma}(\hat{\delta}, \hat{\gamma})$ is the estimate of entity-specific unit standard deviation net of reinsurance, provided by Model M1 and obtained by minimising the criterion function $\ell(\hat{\delta}, \hat{\gamma})$ (specified in Section 1.a) in the interval $D = \{\hat{\delta} \in [0, 1], \hat{\gamma} \in \mathbb{R}\}$. This net coefficient will be obtained either by performing the estimation on net-of-reinsurance data, or by performing the estimation on gross-of-reinsurance data and then multiplying the result by NP_{MW} .

3.c On the minimisation technique

In order to identify $\hat{\delta}$ and $\hat{\gamma}$ one can use either a minimisation routine (an example is the E04JAF routine of the NAG Fortran Library), or a “grid method”, or a combination of the two methods (using the minimum on the grid for initialising the optimisation routine). However the grid computations are useful to analyze the regularity properties of the criterion function.

4 Reserve Risk – Model M1

4.a Specification of the input data

In the *Reserve Risk – Method 1* the data used for estimating the undertaking-specific unit standard deviation of a given segment consists of:

- Y_t : the *year-end obligations* of accounting year t , with $t = 1, \dots, T$ and $T \geq 5$, that is the sum of the payments and the best estimate provisions made at the end of year t for claims occurred in the previous years;
- X_t : the *initial outstanding* of accounting year t , with $t = 1, \dots, T$ and $T \geq 5$, that is the best estimate provisions made at the beginning of year t for claims occurred in the previous years.

It is required that data are representative for the reserve risk which the company will face in the twelve month following the valuation date (i.e. year $t = T + 1$).

Also in this case one could use data concerning direct plus accepted business net of recoveries from policyholders. Since for reserve risk a gross-to-net coefficient is not allowed, this data shall be net of reinsurance recoverables.

The obligations Y_t estimated at the end of the accounting year t are given by:

$$Y_t = P_t + R_t - (P_t^r + R_t^r),$$

where (in brackets it is reported the entry for the gross direct business in the IVASS supervisory form n.17):

- P_t : paid amounts for claims occurred in years previous to accounting year t (v26);
- R_t : claims provision at the end of accounting year t for claims occurred in years previous to accounting year t (v29);
- P_t^r : amounts recovered in accounting year t for deductibles, salvages and subrogations from policyholders and third parties for claims occurred in years previous to accounting year t (v32);
- R_t^r : amounts to be recovered for deductibles, salvages and subrogations from policyholders and third parties at the end of accounting year t for claims occurred in years previous to accounting year t (v33).

The obligations X_t estimated at the beginning of the accounting year t are defined as:

$$X_t = R_{t-1} - R_{t-1}^r + \Delta_t^r + \Delta_t^{cr},$$

where:

- R_{t-1} : claims provision at the end of previous accounting year $t - 1$ (v21);
- R_{t-1}^r : amounts to be recovered from policyholders and third parties at the end of previous accounting year $t - 1$ (v31);
- Δ_t^r : balance of portfolio movements for claims occurred in years previous to accounting year t (v30);
- Δ_t^{cr} : balance of net exchange differences deriving from the updating of foreign currency provisions for claims occurred in years previous to accounting year t (v22).

Remark. In the supervisory forms it is considered the entry relative to the balance of the amounts recovered and to be recovered, that is $R_{t-1}^r - (P_t^r + R_t^r)$. The disaggregated data is present only for the gross direct business. ■

4.b Application of the method

The undertaking-specific standard deviation for segment s according to Method 1 is given by:

$$\sigma_{(res,s,USP)} = c \cdot \hat{\sigma}(\hat{\delta}, \hat{\gamma}) \cdot \sqrt{\frac{T+1}{T-1}} + (1-c) \cdot \sigma_{(res,s)},$$

where:

- T is the length in years of the yearly time series;
- c is the credibility factor;
- $\sigma_{(res,s)}$ is the market-wide level of the unit standard deviation prescribed by EIOPA (which is just defined net of reinsurance);

- $\hat{\sigma}(\hat{\delta}, \hat{\gamma})$ is the estimate of entity-specific unit standard deviation net of reinsurance, provided by Model M1 and obtained by minimising the criterion function $\ell(\hat{\delta}, \hat{\gamma})$ in the interval $D = \{\hat{\delta} \in [0, 1], \hat{\gamma} \in \mathbb{R}\}$ (using, as pointed out, net-of-reinsurance data).

4.c On the minimisation technique

The same arguments apply as in section 3.c.

5 Reserve Risk – Model M2

5.a Specification of the input data

In the *Reserve Risk – Method 2* the data used for estimating the undertaking-specific unit standard deviation of a given segment consists of:

- $X_{i,j}$: amounts for claims occurred in accident year i , with $i = 1, \dots, I$ and $I \geq 5$, and paid with j years of delay, with $j = 1, \dots, J$ and $J \leq I$ (*paid trapezoid*).

The paid amounts $X_{i,j}$ are defined as:

$$X_{i,j} = X_{i,j}^{gr} - X_{i,j}^r,$$

where:

- $X_{i,j}^{gr}$: amounts for claims occurred in accident year i and paid with j years of delay, gross of recovered amounts;
- $X_{i,j}^r$: amounts recovered for deductibles, salvages and subrogations from policyholders and third parties for claims occurred in accident year i and received with j years of delay.

As for Method 1, since a standard gross-to-net adjustment coefficient is not allowed, this data shall be net of reinsurance. With Method 2, however, this requirement is not easily to be fulfilled since “paid-losses triangles” are usually available gross of reinsurance and a gross-to-net transformation can be problematic. Delegated Acts, in D(2)(f) of Annex XVII, prescribe that cumulative payments are adjusted for amounts recoverable from reinsurance contracts which are consistent with the reinsurance contracts *that are in place to provide cover for the following twelve months*. In many cases, however, these adjustments require non-trivial interventions of interpretation and reconstruction which could lead to important distortions of the intrinsic variability of data (that is just what should be estimated).

5.b Application of the method

The undertaking-specific standard deviation for segment s , according to Method 2, is given by:

$$\sigma_{(res,s,USP)} = c \cdot \widehat{\mathbf{Cv}}_{(res,s)} + (1 - c) \cdot \sigma_{(res,s)},$$

where:

- $\widehat{\mathbf{Cv}}_{(res,s)}$ is the estimate of the variation coefficient of the *Outstanding Loss Liabilities* (i.e. the entity-specific unit standard deviation with net-of-reinsurance data) given by (1.19);
- c is the credibility factor;
- $\sigma_{(res,s)}$ is the market-wide level of the unit standard deviation prescribed by EIOPA;

$\widehat{\text{MSEP}}$ and \widehat{R} , the numerator and denominator, respectively, of $\widehat{\mathbf{Cv}}_{(res,s)}$, are derived by the closed-form expressions provided by Model M2 (Merz and Wüthrich model).

Appendix

With regards to the test M1M on the mean and the test M1V on the variance for Model M1 considered in Sections 2.a.1 and 2.a.2, it is worth to make some additional remarks.

A Autocorrelation and heteroscedasticity

If one considers individual data of the type (2.20), that is the time series:

$$(\mathbf{X}, \mathbf{Y}) = \{(X_t, Y_t); t = 1, 2, \dots, T\},$$

since the (X_t, Y_t) are repeated observations for the same company, *autocorrelation* (or *serial correlation*) can be present in the data¹⁷. On the other hand, if one considers market data of the type (2.21), i.e. the time series:

$$\{(\mathbf{X}, \mathbf{Y})_i; i = 1, 2, \dots, N\} = \{(X_{t,i}, Y_{t,i}); i = 1, 2, \dots, N, t = 1, 2, \dots, T_i\},$$

¹⁷The simplest method for detecting the presence of serial correlation is the Durbin-Watson test which provides an estimate of the first-order autocorrelation (i.e. correlation between consecutive residuals). If the test statistic DW (which takes values between 0 and 4) is equal to 2 there are not indications of (first-order) autocorrelation. Values smaller (larger) than 2 indicate positive (negative) autocorrelation. In this framework, a small value of $\mathbf{P}(DW < dw)$ is associated to a high confidence level in positive correlation; a small value of $\mathbf{P}(DW > dw)$ is associated to a high confidence level in negative correlation.

as this data comes from different companies, the presence is possible of *heteroscedasticity* (that is the ε_i variables in (2.22) or in (2.23) do not have the same variance). Thus, in general, the i.i.d. assumptions typical of the standard regression model can fail to be verified. As concerning hypothesis testing, this facts have the important consequence that the values obtained for the estimation errors can be incorrect and then the significance assessments for the parameter estimates can be not reliable. In particular, if there is positive autocorrelation in the data, the regression residuals can have lower dispersion which implies that we can underestimate the estimation error and, consequently, we can overestimate the significance of the corresponding parameter.

To overcome these difficulties the assessment of the estimation errors must be properly corrected. This can be done by performing the estimates in the form which is referred to as HAC (*Heteroscedasticity and Autocorrelation Consistent*), basically consisting in deriving estimates for the model covariance matrix which include corrections for the effects of heteroscedasticity and autocorrelation possibly present in the data. For a review on the main HAC estimators one can see [25] and [2].

B Resampling of individual data

Another issue concerns the possibility of using individual (i.e. entity-specific) data for the M1M and M1V tests. At least in principle, a possible way to avoid the use of market data, particularly for the M1V test on the variance, consists in *resampling* the individual time series (\mathbf{X}, \mathbf{Y}) by a *bootstrap* procedure. If the observed time series:

$$(\mathbf{X}, \mathbf{Y}) = \{(X_t, Y_t); t = 1, 2, \dots, T\},$$

is a realization of i.i.d. variables, the standard (non-parametric) bootstrap consists in producing a set of n “pseudoserries”:

$$\left\{ (\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})_{\kappa}; \kappa = 1, 2, \dots, n \right\}, \quad (\text{B.39})$$

where the κ -th pseudoserries $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})_{\kappa}$ is obtained by making T equiprobable sampling with replacement from the original time series (\mathbf{X}, \mathbf{Y}) . Then one can perform the M1M and M1V test on the n “observations” of the mean and the variance, respectively, obtained by the pseudoserries (B.39).

In order that the bootstrap approach provides reliable results, however, the statistical properties of the resampled series (B.39) must be as close as possible to that of the original series. This can be a problem if there is autocorrelation in the original data because in the standard bootstrap approach the single observations, as being randomly sampled, are also randomly re-sorted, which destroys the dependence structure possibly present

in the data. A way to overcome this problem could be the use of *block bootstrap* methods, where the pseudoseries $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})_\kappa$ are obtained by resampling blocks of consecutive observations, instead of single observations. This should allow to capture, at least, the dependence structure internal to the blocks.

Block bootstrap can be realized in many different ways. The most appropriate to the applications we are interested in here seems to be the *Moving Block Bootstrap* (MBB), which uses overlapping blocks with fixed length (see [18] [20]). Let $l < T$ be the block length. On the original series one defines, by a “moving-window” approach, $T - l + 1$ blocks:

$$\begin{aligned} &\{(X_1, Y_1), (X_2, Y_2), \dots, (X_l, Y_l)\}, \\ &\{(X_2, Y_2), (X_3, Y_3), \dots, (X_{l+1}, Y_{l+1})\}, \\ &\dots \\ &\{(X_{T-l+1}, Y_{T-l+1}), (X_{T-l+2}, Y_{T-l+2}), \dots, (X_T, Y_T)\}. \end{aligned}$$

Then one performs the κ -th resampling by randomly drawing with replacement from this set $\lceil T/l \rceil$ blocks¹⁸ and then collecting these blocks in the order of drawing until a pseudoseries $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})_\kappa$ with length T is obtained (a final truncation is made when necessary).

The accuracy of the block bootstrap in replicating the data is sensitive to the choice of the block length l , and so far generally accepted criteria are not available for defining the optimal value of this parameter. As a rule of thumb one could pose $l \approx T^{1/3}$ but more sophisticated criteria are available, that are chosen considering the properties assumed for the data and the specific kind of application (see e.g. [19]). Given the short length T of the time series available in USP applications, the choice of sensible values for the block length is very limited. Practically, it seems reasonable to consider for l the values 2, 3 or 4.

¹⁸We denote by $\lceil x \rceil$ the minimum integer greater than or equal to x .

References

- [1] Anderson, T. W., Darling, D. A., *A test of goodness of fit*, Journal of the American statistical Association, 765-769, 1954.
- [2] Andrews D.W.K., *Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation*, Econometrica **59**, 817-858, 1991.
- [3] Cramér, H., *On the composition of elementary errors*, Scandinavian Actuarial Journal, 1928.
- [4] Buchwalder, M., Bühlmann, H., Merz, M., Wüthrich, M.V., *The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited)*, ASTIN Bulletin **36(2)**, 521-542, 2006.
- [5] De Felice, M., Moriconi, F., *Risk based capital in P&C loss reserving or stressing the triangle*, Research Group on “Insurance Companies and Pension Funds”, Working Paper n.1, Rome, December 2003.
- [6] De Felice, M., Moriconi, F., *Process error and estimation error of year-end reserve estimation in the distribution free chain-ladder model*, Alef Working Paper, Rome, November 2006.
- [7] De Felice, M., Moriconi, F., *Measuring Goodness of Fit of Loss Reserving Models*, Rome, Alef Working Paper, July 2008.
- [8] De Felice, M., Moriconi, F., Cavastracci, S., Matarazzo, L., Pasqualini, S., *Reserve requirements and capital requirements in non-life insurance. An analysis of the Italian MTPL insurance market by stochastic claims reserving models*, ISVAP, Rome, October 2006.
- [9] EIOPA, *Calibration of the Premium and Reserve Risk factors in Standard Formula of Solvency II*, Report of the Joint Working Group on Non-Life and Health NSLT Calibration, 12 December 2011.
- [10] Embrechts, P., Klüppelberg, C., Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag 1997.
- [11] England, P. , *Addendum to “Analytic and bootstrap estimates of prediction errors in claims reserving”*, Insurance: Mathematics and Economics, vol. 31, pp. 461-466, 2002.
- [12] England, P., Verral, R., *Analytic and bootstrap estimates of prediction errors in claims reserving*, Insurance: Mathematics and Economics, vol. 25 pp. 281-293, 1999.
- [13] European Commission, *Text of the delegated Acts on Solvency II*, 10 October 2014.
(<http://ec.europa.eu/finance/insurance/solvency/solvency2/index.en.htm>)
- [14] European Commission, *Annex 1 to 21 of the Delegated Acts on Solvency II*, 10 October 2014.
(<http://ec.europa.eu/finance/insurance/solvency/solvency2/index.en.htm>)

- [15] IVASS, *Solvency II – utilizzo dei parametri specifici dell’impresa nei moduli del rischio di sottoscrizione per l’assicurazione vita, per l’assicurazione non vita e per l’assicurazione malattia*, Lettera al mercato, 4 novembre 2014.
- [16] Jarque, C.M., Bera, A.K., *Efficient tests for normality, homoscedasticity and serial independence of regression residuals*, Economics Letters, **6**(3), 255-259, 1980.
- [17] Klugman, S.A., Panjer, H.H., Willmot, E., *Loss Models. From data to decisions*, Wiley 2012.
- [18] Künsch, H.R., *The jackknife and the bootstrap for general stationary observations*, Ann. Statist. **17**, 1217-1241, 1989.
- [19] Lahiri, S.N., Furukawa, K., Lee, Y.D., *A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods*, Statistical Methodology **4**, 292-321 2007.
- [20] Liu, R.Y., Singh, K., *Moving blocks jackknife and bootstrap capture weak dependence*. In *Exploring the Limits of Bootstrap* (LePage R., Billard L. eds.), Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. 225-248, Wiley 1992.
- [21] Mack, T. , *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates* ASTIN Bulletin **23**(2), 213-225, 1993.
- [22] Massey, Jr., Frank J., *The Kolmogorov-Smirnov test for goodness of fit*, Journal of the American Statistical Association, 68-78, 1951.
- [23] Merz, M., Wüthrich, M. V., *Modelling Claims Development Result for Solvency Purposes*, Casualty Actuarial Society, E-Forum, Fall 2008.
- [24] Mikosch, T., *Non-Life Insurance Mathematics. An introduction with the Poisson process*, 2nd ed., Universitext, Springer, Berlin, 2009.
- [25] Newey, W.K, West, K., *A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, Econometrica **55**, 703-708, 1987.
- [26] Shapiro, S.S, Wilk, M.B., *An analysis of variance test for normality (complete samples)*, Biometrika **52**(3-4), 591-611, 1965
- [27] Venter, G., *Testing the assumptions of age to age factors*, PCAS 1998.
- [28] Verral, R.J. *An investigation into stochastic claims reserving models and the chain-ladder technique*, Insurance: Mathematics and Economics, vol. 26, pp. 91-99, 2000.
- [29] Verrall, R.J., England, P.D., *Comments on “A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates”, by Mack and Venter*, Insurance: Mathematics and Economics, vol. 26, pp. 109-111, 2000.
- [30] von Mises, R.E., *Wahrscheinlichkeit, Statistik und Wahrheit*, Julius Springer, 1928.

- [31] Wüthrich, M. V., Merz, M., *Stochastic Claims Reserving Methods in Insurance*, Wiley Finance, 2008.