Scandinavian Actuarial Journal

ISSN: 0346-1238 (Print) 1651-2030 (Online) Journal homepage: https://www.tandfonline.com/loi/sact20

Generalized log-normal chain-ladder

D. Kuang & B. Nielsen

To cite this article: D. Kuang & B. Nielsen (2019): Generalized log-normal chain-ladder, Scandinavian Actuarial Journal, DOI: 10.1080/03461238.2019.1696885

To link to this article: https://doi.org/10.1080/03461238.2019.1696885

SCANDINAVIAN ACTUARIAL JOURNAL
https://doi.org/10.1080/03461238.2019.1696885

Generalized log-normal chain-ladder

D. Kuang[a] and B. Nielsen[b,c]

[a]Market Reserve & Capital, Lloyd's of London, London UK; [b]Nuffield College, Oxford, UK; [c]Department of Economics, University of Oxford, Oxford, UK

ABSTRACT

We propose an asymptotic theory for distribution forecasting from the log-normal chain-ladder model. The theory overcomes the difficulty of convoluting log-normal variables and takes estimation error into account. The results differ from that of the over-dispersed Poisson model and from the chain-ladder-based bootstrap. We embed the log-normal chain-ladder model in a class of infinitely divisible distributions called the generalized log-normal chain-ladder model. The asymptotic theory uses small σ asymptotics where the dimension of the reserving triangle is kept fixed while the standard deviation is assumed to decrease. The resulting asymptotic forecast distributions follow t distributions. The theory is supported by simulations and an empirical application.

ARTICLE HISTORY

Received 24 October 2019
Accepted 20 November 2019

KEYWORDS

Chain-ladder; infinitely divisibility; over-dispersed Poisson; bootstrap; log-normal

1. Introduction

Reserving in general insurance usually relies on chain-ladder-type methods. The most popular method is the traditional chain-ladder. A contender is the log-normal chain-ladder, which we study here. Both methods have proved to be valuable for point forecasting. In practice, distribution forecasting is needed too. For the standard chain-ladder, there are presently three methods available. Mack (1999) has suggested a method for recursive calculation of standard errors of the forecasts, but without proposing an actual forecast distribution. The bootstrap method of England & Verrall (1999) and England (2002) is commonly used, but it does not always produce satisfactory results. Recently, Harnau & Nielsen (2018) have developed an asymptotic theory for the chain-ladder in which the idea of a over-dispersed Poisson framework is embedded in a formal model. This was done through a class of infinitely divisible distributions and a new Central Limit Theorem. An asymptotic theory provides an analytic tool for evaluating the distribution of forecast errors and building inferential procedures and specification tests for the model. Here we adapt the infinitely divisible framework to the log-normal chain-ladder and present an asymptotic theory for the distribution forecasts and model evaluation. Thereby, asymptotic distribution forecasts and model evaluation tools are now available for two different models, which together cover a wide range of reserving triangles.

The data consist of a reserving triangle of aggregate amounts that have been paid with some delay in respect to portfolios of insurances. Table 1 provides an example. The objective of reserving is to forecast liabilities that have occurred but have not yet been settled or even recorded. The reserve is an estimate of these liabilities. Thus, the problem is to forecast the lower reserving triangle and then add these forecasts up to get the reserve. The traditional chain-ladder provides a point forecast for the reserve.

CONTACT B. Nielsen bent.nielsen@nuffield.ox.ac.uk Nuffield College, Oxford OX1 1NF, UK.

Supplemental data for this article can be accessed here. https://doi.org/10.1080/03461238.2019.1696885

Table 1. XL Group, US casualty, gross paid and reported loss and allocated loss adjustment expense in 1000 USD.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1997	2185	13,908	44,704	56,445	67,313	62,830	72,619	42,511	32,246	51,257	11,774	21,726	10,926	4763	3580	4777	1070	1807	824	1288
1998	3004	17,478	49,564	55,090	75,119	66,759	76,212	62,311	31,510	15,483	23,970	8321	15,027	3247	8756	14,364	3967	3858	4643	
1999	5690	28,971	55,352	63,830	71,528	73,549	72,159	37,275	38,797	27,264	28,651	14,102	8061	17,292	10,850	10,732	4611	4608		
2000	9035	29,666	47,086	41,100	58,533	80,538	70,521	40,192	27,613	13,791	17,738	20,259	12,123	6473	3922	3825	3082			
2001	7924	38,961	41,069	64,760	64,069	61,135	62,109	52,702	36,100	18,648	32,572	17,751	18,347	10,895	2974	5828				
2002	7285	25,867	44,375	58,199	61,245	48,661	57,238	29,667	34,557	8560	12,604	8683	9660	4687	1889					
2003	3017	22,966	62,909	54,143	72,216	58,050	29,522	25,245	19,974	16,039	8083	9594	3291	2016						
2004	1752	25,338	56,419	75,381	64,677	58,121	38,339	21,342	14,446	13,459	6364	6326	6185							
2005	1181	24,571	66,321	65,515	62,151	43,727	29,785	23,981	12,365	12,704	12,451	8272								
2006	1706	13,203	40,759	57,844	48,205	50,461	27,801	21,222	14,449	10,876	8979									
2007	623	14,485	27,715	52,243	60,190	45,100	31,092	22,731	19,950	18,016										
2008	338	6254	24,473	32314	35698	25849	30407	15335	15697											
2009	255	3842	14,086	26,177	27,713	15,087	17,085	12,520												
2010	258	7426	22,459	28,665	32,847	28,479	24,096													
2011	1139	10,300	19,750	32,722	41,701	29,904														
2012	381	5671	34,139	33,735	33,191															
2013	605	11,242	24,025	32,777																
2014	1091	9970	31,410																	
2015	1221	8374																		
2016	2458																			

The chain-ladder is maximum likelihood in a Poisson model. This is useful for estimation and point forecasting. Martínez-Miranda et al. (2015) have developed a theory for inference and distribution forecasting in such a Poisson model in order to analyse and forecast incidences of mesothelioma. However, this is not of much use for the reserving problem because the data is nearly always severely over-dispersed. The over-dispersion arises because each entry in the paid triangle is the aggregate amount paid out to an unknown number of claims of different severity. It is common to interpret this as a compound Poisson variable, see Beard et al. (1984, Section 3.2). Compound Poisson variables are indeed over-dispersed in the sense that the variance to mean ratio is larger than unity. They are, however, difficult to analyse and even harder to convolute. England & Verrall (1999) and England (2002) developed a bootstrap to address this issue. This often works, but it is known to give unsatisfactory results in some situations. The model underlying the bootstrap is not fully described, so it is hard to show formally when the bootstrap is valid and to generalize it to other situations, including the log-normal chain-ladder.

The infinitely divisible framework of Harnau & Nielsen (2018) provides a plausible over-dispersed Poisson model and framework for distribution forecasting with the traditional chain-ladder. It utilizes that the compound Poisson distribution is infinitely divisible. If the mean of each entry in the paid triangle is large, then the skewness of compound Poisson variable is small and a Central Limit Theorem applies. Thus, keeping the dimension of the triangle fixed, while letting the mean increase, the reserving triangle is asymptotically normal with mean and variance estimated by the chain-ladder. Since the dimension is fixed we then arrive at an asymptotic theory that matches the traditional theory for analysis of variance (anova) developed by Fisher in the 1920s. If the over-dispersion is unity and therefore known as in the Poisson model of Martínez-Miranda et al. (2015) then inference is asymptotically χ^2 and distribution forecasts are normal. When the over-dispersion is estimated as appropriate for reserving data then we arrive at inference that is asymptotically F and distribution forecasts that are asymptotically t. The chain-ladder bootstrap could potentially be analysed within this framework, but this is yet to be done.

When it comes to the log-normal model the situation is different. The log-normal model has apparently been suggested by Taylor in 1979, and then analysed by for instance Kremer (1982), Renshaw (1989), Verrall (1991, 1994), Doray (1996) and England & Verrall (2002). The main difference to the over-dispersed Poisson model is that the mean-variance ratio is constant across the triangle in that model, while the mean-standard deviation ratio is constant in the log-normal model. Therefore, the tails of distributions are expected to be different, which may matter in distribution forecasting.

Estimation is easy in the log-normal model. It is done by least squares from the log triangle. Recently, Kuang et al. (2015) have provided exact expressions for all estimators along with a set of associated development factors. Least squares theory provides a distribution theory for the estimators and for inference. However, the reserving problem is to make forecasts of reserves that are measured on the original scale. Each entry in the original scale is log-normally distributed. While there are expressions for such log-normal distributions it is unclear how to incorporate estimation uncertainty, let alone convolute such variables to get the reserve.

The infinitely divisible theory provides a solution also for the log-normal model. Thorin (1977) showed that the log-normal distribution is infinitely divisible. First of all, this indicates that the log-normal variables actually have an interpretation as compound sums of claims. Secondly, the framework of Harnau & Nielsen (2018) and their Central Limit Theorem apply, albeit with subtle differences. In the over-dispersed Poisson model the mean of each entry is taken to be large in the asymptotic theory, whereas for generalized log-normal model we will let the variance be small in the asymptotic theory. In both cases, the mean-dispersion ratio is then large. In this paper we will exploit that infinitely divisible theory to provide an asymptotic theory for the log-normal distribution forecasts.

We also discuss specification tests for the log-normal model. Mis-specification can appear both in the mean and the variance of the log-normal variables. The mean could for instance have an omitted calendar effect. Thus, we study the extended chain-ladder model discussed by Zehnwirth (1994), Barnett & Zehnwirth (2000), and Kuang et al. (2008a, 2008b, 2011). The variance could be different in subgroups of the triangle as pointed out by Hertig (1985). Bartlett (1937) proposed a test for this problem. Recently, Harnau (2018a) has adapted that test to the traditional chain-ladder. We extend this to the generalized log-normal model.

We illustrate the new methods using a casualty reserving triangle from XL Group (2017) as shown in Table 1. The triangle is for US casualty and includes gross paid and reported loss and allocated loss adjustment expense in 1000 USD.

We conduct a simulation study where the data generating process matches the XL Group data in Table 1. We find that that the asymptotic results give good approximations in finite samples. The asymptotics will work even better if the mean-dispersion ratio is larger. The generalized log-normal model is also compared with the over-dispersed Poisson model and the England (2002) bootstrap. The bootstrap is found not to work very well by an order of magnitude for this log-normal data generating process. The over-dispersed Poisson model works better although it is dominated by the generalized log-normal model.

In Section 2, we review the well-known log-normal models for reserving. In Section 3, we set up the asymptotic generalized log-normal model based on the infinitely divisible framework. We check that the log-normal model is embedded in this class and show that the results for inference in the log-normal model carries over to the generalized log-normal model. We also derive distribution forecasts. We apply the results to the XL Group data in Section 4, while Section 5 provides the simulation study. Finally, we discuss directions for future research in Section 6. All proofs of theorems are provided in an Appendix.

2. Review of the log-normal chain-ladder model

A competitor to the chain-ladder is the log-normal model. In this model, the log of the data is normal so that parameters can be estimated by ordinary least squares. We review the log-normal model by describing the structure of the data, the model, statistical analysis, point forecasts and extension by a calendar effect.

2.1. Data

Consider a standard incremental insurance run-off triangle of dimension k. Each entry is denoted Y_{ij} so that i is the origin year, which can be accident year, policy year or year of account, while j is the development year. Collectively, we have data $\mathbf{Y} = \{Y_{ij}, \forall(i,j) \in \mathcal{I}\}$, where $\mathcal{I}$ is the triangular index set

$$\mathcal{I} = \{(i,j) : i \text{ and } j \text{ belong to } (1,\dots,k) \text{ with } i+j-1 = 1,\dots,k\}. \tag{1}$$

Let $n = k(k+1)/2$ be the number of observations in the triangle $\mathcal{I}$. One could allow more general index sets, see Kuang et al. (2008a), for instance to allow for situations where some accidents are fully run-off or only recent calendar years are available. We are interested in forecasting the lower triangle with index set

$$\mathcal{J} = \{(i,j) : i \text{ and } j \text{ belong to } (1,\dots,k) \text{ with } i+j-1 = k+1,\dots,2k-1\}. \tag{2}$$

2.2. Model

In the log-normal model, the log claims have expectation given by the linear predictor

$$\mu_{ij} = \alpha_i + \beta_j + \delta. \tag{3}$$

The predictor μ_{ij} is composed of a an accident year effect α_i, a development year effect β_j and an overall level δ. The model is then defined as follows.

Assumption 2.1 (log-normal model): *The array* $\mathbf{Y} = \{Y_{ij}, \forall(i,j) \in \mathcal{I}\}$, *satisfies that the variables* $y_{ij} = \log Y_{ij}$ *are independent normal* $\mathsf{N}(\mu_{ij}, \omega^2)$ *distributed, where the mean* μ_{ij} *is given by* (3).

The parametrization presented in (3) does not identify the distribution. It is common to identify the parameters by setting, for instance, $\delta = 0$ and $\sum_{j=1}^{k} \beta_j = 0$. Such an ad hoc identification makes it difficult to extrapolate the model beyond the square composed of the upper triangle $\mathcal{I}$ and the lower triangle $\mathcal{J}$ and it is not amenable to the subsequent asymptotic analysis. Thus, we switch to the canonical parametrization of Kuang et al. (2009, 2015) so that the model becomes a regular exponential family with freely varying parameters. The canonical parameter is

$$\xi = \{\mu_{11}, \Delta\alpha_2, \dots, \Delta\alpha_k, \Delta\beta_2, \dots, \Delta\beta_k\}', \tag{4}$$

where $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$ is the relative accident year effect and $\Delta\beta_j = \beta_j - \beta_{j-1}$ is the relative development year effect, while μ_{11} is the overall level. The length of ξ is denoted p, which is $p = 2k - 1$ with the chain-ladder structure. We can then write

$$\mu_{ij} = \mu_{11} + \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell = X_{ij}'\xi, \tag{5}$$

with the convention that empty sums are zero and $X_{ij} \in \mathbb{R}^p$ is the design vector

$$X_{ij}' = \{1, 1_{(2\le i)}, \dots, 1_{(k\le i)}, 1_{(2\le j)}, \dots, 1_{(k\le j)}\}, \tag{6}$$

where the indicator function $1_{(m\le i)}$ is unity if $m \le i$ and zero otherwise.

2.3. Statistical analysis

The log observations $y_{ij} = \log Y_{ij}$ have a normal log likelihood given by

$$\ell_{\log \mathsf{N}}(\xi, \omega^2) = -\frac{n}{2}\log(2\pi\omega^2) - \frac{1}{2\omega^2}\sum_{(i,j)\in\mathcal{I}} (y_{ij} - X_{ij}'\xi)^2. \tag{7}$$

Stacking the observations $y_{ij} = \log Y_{ij}$ and the row vectors X_{ij}' then gives an observation vector y and a design matrix X and a model equation of the form

$$y = X\xi + \varepsilon. \tag{8}$$

The least squares estimator for ξ and the residuals are then

$$\hat{\xi} = (X'X)^{-1}X'y, \quad \hat{\varepsilon}_{ij} = y_{ij} - X_{ij}'\hat{\xi}. \tag{9}$$

while the variance ω^2 is estimated by

$$s^2 = \frac{RSS}{n-p} \quad \text{where} \quad RSS = \sum_{(i,j)\in\mathcal{I}} \hat{\varepsilon}_{ij}^2. \tag{10}$$

Kuang et al. (2015) derive explicit expressions for each coordinate of the canonical parameter and they provide an interpretation in terms of so-called geometric development factors.

Standard least squares theory provides a distribution theory for the estimators, see for instance Hendry & Nielsen (2007), so that

$$\hat{\xi} \overset{\mathsf{D}}{=} \mathsf{N}\{\xi, \omega^2(X'X)^{-1}\}, \quad s^2 \overset{\mathsf{D}}{=} \chi^2_{n-p}/(n-p). \tag{11}$$

Individual components of $\hat{\xi}$ will also be normal. Standardizing those components and replacing ω^2 by the estimate s^2 gives the t -statistic, which is t_{n-p} distributed.

We may be interested in testing linear restrictions on ξ. This can be done using F-tests. For instance, the hypothesis that all $\Delta\alpha$ parameters are zero would indicate that the policy year effect is constant over time. Such restrictions can be formulated as $\xi = H\zeta$ for some known matrix $H \in \mathbb{R}^{p\times p_H}$ and a parameter vector $\zeta \in \mathbb{R}^{p_H}$. In the example of zero $\Delta\alpha$'s the H matrix would select the remaining parameters, the μ_{11} and the $\Delta\beta_j$s. We then get a restricted design matrix $X_H = XH$ and a model equation of the form $y = X_H\zeta + \varepsilon$. We then get estimators

$$\hat{\zeta} = (X_H'X_H)^{-1}X_H'y, \quad s_H^2 = \frac{RSS_H}{n - p_H},$$

where the residual sum of squares $RSS_H = \sum_{(i,j)\in\mathcal{I}} \hat{\varepsilon}_{H,ij}^2$ is formed from the residuals $\hat{\varepsilon}_{H,ij} = y_{ij} - X_{H,ij}'\hat{\zeta}$ as before. The hypothesis can be tested by F-statistic

$$F = \frac{\{RSS_H - RSS\}/(p - p_H)}{RSS/(n-p)} \overset{\mathsf{D}}{=} \mathsf{F}(p - p_H, n - p_H). \quad (12)$$

We may also be interested in affine restrictions. For instance, the hypothesis that all $\Delta\alpha$ parameters are known corresponds the hypothesis of known values of relative ultimates. This may be of interest in an Bornhuetter-Ferguson context, see Elpidorou et al. (2019). This is analysed by restricted least squares which also leads to t and F statistics.

2.4. Point forecasting

In practice, we will want to forecast the variables Y_{ij} on the original scale. Since y_{ij} is $\mathsf{N}(\mu_{ij}, \omega^2)$ then $Y_{ij} = \exp(y_{ij})$ is log-normally distributed with mean $\exp(\mu_{ij} + \omega^2/2)$. Thus, the point forecast for the lower triangle $\mathcal{J}$ can be formed as

$$\tilde{Y}_{ij} = \exp(X_{ij}'\hat{\xi} + \hat{\omega}^2/2), \quad (13)$$

where the tilde indicate that $\tilde{Y}_{ij}$ is an out-of-sample forecast. When it comes to distribution forecasting, the log-normal model has the drawback that it is a non-trivial problem to characterize the joint distribution of the variables on the original scale. Renshaw (1989) provides expressions for the covariance matrix of the variables on the original scale, but a further non-trivial step would be needed to characterize the joint distribution. Once it comes to distribution forecasting we would also need to take the estimation error into account. This does not make the problem easier. We will circumvent these issues by exploiting the infinitely divisible setup of Harnau & Nielsen (2018).

2.5. Extending with a calendar effect

It is common to extend the chain-ladder parametrization with a calendar effect, so that linear predictor in (3) becomes

$$\mu_{ij,apc} = \alpha_i + \beta_j + \gamma_{i+j-1} + \delta, \quad (14)$$

where $i+j-1$ is the calendar year corresponding to accident year i and development year j. This model has been suggested in insurance by Zehnwirth (1994). Similar models have been used in a variety of disciplines under the name of age-period-cohort models, where age, period and cohort are our development, calendar and policy year. The model has an identification problem. The canonical parameter solution of Kuang et al. (2008a) is to write $\mu_{ij,apc} = X_{ij,apc}'\xi_{apc}$ where, with $h(i,s) = \max(i - s + 1, 0)$, we have

$$\xi_{apc} = (\mu_{11}, \nu_a, \nu_c, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_k, \Delta^2\beta_3, \ldots, \Delta^2\beta_k, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_k)', \quad (15)$$

$$X_{ij,apc} = \{1, i-1, j-1, h(i,3), \dots, h(i,k), h(j,3), \dots, h(j,k),$$
$$h(i+j-1,3), \dots, h(i+j-1,k)\}. \tag{16}$$

The dimension of these vectors is $p_{apc} = 3k - 3$.

This model can be analysed by the same methods as above. Stack the design vectors $X'_{ij,apc}$ to a design matrix X_{apc} and regress y on X_{apc} to get an estimator ξ_{apc} of the form (9) along with a residual sum of squares RSS_{apc} and a variance estimator s^2_{apc} The significance of the calendar effect can be tested using an F-statistic as in (12), where ξ and p now correspond to the extended model, while ζ and p_H correspond to the chain-ladder specification.

When it comes to forecasting it is necessary to extrapolate the calendar effect. This has to be done with some care due to an identification problem, see Kuang et al. (2008b, 2011).

3. The generalized log-normal chain-ladder model

The log-normal distribution is infinitely divisible as shown by Thorin (1977). We can therefore formulate a class of infinitely divisible distributions encompassing the log-normal. We will refer to this class of distributions as the generalized log-normal chain-ladder model. In the analysis we exploit the setup of Harnau & Nielsen (2018) to provide distribution forecasts for the generalized log-normal model.

3.1. Assumptions and first properties

The infinitely divisible setup of Harnau & Nielsen (2018, Section 3.7) encompasses the log-normal model. Recall that a distribution D is infinitely divisible, if for any $m \in \mathbb{N}$, there are independent, identically distributed random variables $X_1, \dots, X_m$ such that $\sum_{\ell=1}^{m} X_\ell$ has distribution D. The log-normal distribution is infinitely divisible as shown by Thorin (1977). This matches the fact that the paid amounts are aggregates of number of payments. In our data analysis we neither know the number nor the severities of the payments. Due to the infinite divisibility the log-normal distribution can therefore be a good choice for modelling aggregate payments.

We will propose a small variance asymptotic theory. In standard asymptotic theory for repeated samples the asymptotics theory applies as the sample size increases. In that case a consequence is that averages have decreasing variance. Here, there is no repeated sampling. Instead, data are constructed by aggregation over many policies and assumptions are formulated directly in terms of decreasing variances. The infinite divisible construction implies that the repetitive structure is implicit. We will formulate two assumptions. The first assumption is about a general infinite divisible setup. The second assumption gives more specific details on the log-normal setup.

Assumption 3.1 (Infinite divisibility): *The array Y_{ij}, $(i,j) \in \mathcal{I}$, satisfies*

(i) *Y_{ij} are independent distributed, non-negative and infinitely divisible*;
(ii) *the dimension of the array $\mathcal{I}$ is fixed*;
(iii) *asymptotically, the skewness vanishes*: $\mathsf{skew}(Y_{ij}) = \mathsf{E}[\{Y_{ij} - \mathsf{E}(Y_{ij})\}/\mathsf{sdv}(Y_{ij})]^3 \to 0$.

We have the following Central Limit Theorem for non-negative, infinitely divisible distributions with vanishing skewness. This is different from the standard Lindeberg-Lévy Central Limit Theorem for averages of independent, identically distributed variables, but proved in a similar fashion by analysing characteristic function and exploiting the Lévy-Kintchine formula for infinitely divisible distributions.

Theorem 3.1 (Harnau & Nielsen 2018, Theorem 1): *Suppose Assumption 3.1 is satisfied. Then, as the skewness vanishes, the sequence of standardized entries Y_{ij} converge in distribution to normal* *distributions, that are independent over* $(i, j) \in \mathcal{I}$*. That is,*

$$\frac{Y_{ij} - \mathsf{E}(Y_{ij})}{\sqrt{\mathsf{Var}(Y_{ij})}} \overset{\mathsf{D}}{\to} \mathsf{N}(0, 1).$$

We need some more specific assumptions for the log-normal setup. When describing the predictor we write $\mu_{ij} = X'_{ij}\xi$ to indicate that any linear structure is allowed as long as ξ is freely varying when estimating in the statistical model. This could be the chain-ladder structure as in (5), (6) or an extended chain-ladder model with a calendar effect.

Assumption 3.2 (The generalized log-normal chain-ladder model): *The array* Y_{ij}, $(i, j) \in \mathcal{I}$, *satisfies Assumptions* 3.1 *and the following:*

(i) $\log \mathsf{E}(Y_{ij}) = \mu_{ij} + \omega^2/2 = X'_{ij}\xi + \omega^2/2$, *where* ξ *is identified by the likelihood* (7);
(ii) *asymptotically,* $\omega^2 \to 0$ *while* ξ *is fixed;*
(iii) *asymptotically,* $\mathsf{Var}(Y_{ij})/\{\omega^2\mathsf{E}^2(Y_{ij})\} \to 1$.

We check that the log-normal model set out in Assumption 2.1 is a generalized log-normal model.

Theorem 3.2: *Consider the log-normal model of Assumption* 2.1*. Suppose the dimension of the array* $\mathcal{I}$ *is fixed as* $\omega^2 \to 0$*. Then Assumptions* 3.1 *and* 3.2 *are satisfied.*

A first consequence of the generalized log-normal model is that Theorem 3.1 provides an asymptotic theory for the claims on the original scale. We now check that we have a normal theory for the log claims. The proof applies the delta method. Theorem 3.3 is useful in deriving the inference in Theorem 3.5 and estimation error for forecasts in Theorem 3.8 in later sections.

Theorem 3.3: *Suppose Assumptions* 3.1 *and* 3.2 *are satisfied. Let* $y_{ij} = \log Y_{ij}$*. Then, as* $\omega^2 \to 0$,

$$\omega^{-1}(y_{ij} - \mu_{ij}) \overset{\mathsf{D}}{\to} \mathsf{N}(0, 1).$$

Due to the independence of Y_{ij} *over* $(i, j) \in \mathcal{I}$ *the standardized* y_{ij} *are asymptotically independent standard normal.*

We will need to reformulate the Central Limit Theorem 3.1 slightly. The issue is that the generalized log-normal model leaves the variance of the variable unspecified in a finite sample, so that the Central Limit Theorem is difficult to manipulate directly. Theorem 3.4 is useful in deriving the process error for forecasts in Theorem 3.8 later.

Theorem 3.4: *Suppose Assumptions* 3.1 *and* 3.2 *are satisfied. Then, as* $\omega^2 \to 0$,

$$\omega^{-1}\{Y_{ij} - \mathsf{E}(Y_{ij})\} \overset{\mathsf{D}}{\to} \mathsf{N}\{0, \exp(2\mu_{ij})\}.$$

3.2. Inference

We check that the inferential results for the log-normal model, described in Section 2.3, carry over to the generalized log-normal model. First, we consider the asymptotic distribution of estimators and then the properties of F-statistics for inference.

Theorem 3.5: *Consider the generalized log-normal model defined by Assumptions* 3.1 *and* 3.2 *and the least squares estimators* (9)*. Then, as* $\omega^2 \to 0$,

$$\omega^{-1}(\hat{\xi} - \xi) \overset{\mathsf{D}}{\to} \mathsf{N}\{0, (X'X)^{-1}\},$$

$$\omega^{-2}s^2 \xrightarrow{\mathsf{D}} \chi^2_{n-p}/(n-p),$$

where $\hat{\xi}, s^2$ *are asymptotically independent.*

We can derive inference for the parameter ξ using an asymptotic t distribution. The proof follows from Theorem 3.5 and the Continuous Mapping Theorem by considering ratios of $\hat{\xi}, s^2$, noting that the two estimators are asymptotically independent.

Theorem 3.6: *Consider the generalized log-normal model, defined by Assumptions* 3.1 *and* 3.2. *Then, as* $\omega^2 \to 0$,

$$\frac{v'(\hat{\xi}-\xi)}{s\sqrt{v'(X'X)^{-1}v}} \xrightarrow{\mathsf{D}} \mathsf{t}_{n-p}$$

We can also make inference using asymptotic F-statistics, mirroring the F-statistic (12) from the classical normal model. The proof is similar to Theorem 4 of Harnau & Nielsen (2018).

Theorem 3.7: *Consider the generalized log-normal model, defined by Assumptions* 3.1 *and* 3.2 *with three types of linear predictor*:

(1) *the extended chain-ladder model parametrized by* $\xi_{apc} \in \mathbb{R}^{p_{apc}}$ *in* (15);
(2) *the chain-ladder model parametrized by* $\xi \in \mathbb{R}^p$ *in* (4); *and*
(3) *a linear hypothesis* $\xi = H\zeta$ *for* $\zeta \in \mathbb{R}^{p_H}$ *and some known matrix* $H \in \mathbb{R}^{p \times p_H}$.

Let RSS_{apc}, RSS *and* RSS_H *be the residual sums of squares under the linear hypotheses. Then, as* $\omega \to 0$,

$$F_1 = \frac{(RSS - RSS_{apc})/(p_{apc}-p)}{RSS_{apc}/(n-p_{apc})} \xrightarrow{\mathsf{D}} \mathsf{F}_{p_{apc}-p,n-p_{apc}},$$
$$F_2 = \frac{(RSS_H - RSS)/(p-p_H)}{RSS/(n-p)} \xrightarrow{\mathsf{D}} \mathsf{F}_{p-p_H,n-p},$$

where F_1 *and* F_2 *are asymptotically independent.*

3.3. Distribution forecasting

The aim is to predict a sum of elements in the lower triangle, that could be the overall sum, which is the total expected reserve; or it could be row sums or diagonal sums giving the cash flow of the expected reserve. We denote such sums by $Y_{\mathcal{A}} = \sum_{(i,j)\in\mathcal{A}} Y_{ij}$ for some subset $\mathcal{A} \in \mathcal{J}$. The point forecasts for a single entry are $\hat{Y}_{ij} = \exp(X'_{ij}\hat{\xi} + s^2/2)$ as given in (13), while the overall point forecast is

$$\tilde{Y}_{\mathcal{A}} = \sum_{(i,j)\in\mathcal{A}} \tilde{Y}_{ij} = \sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\hat{\xi} + s^2/2) \tag{17}$$

To find the forecast error, we expand

$$\begin{aligned} Y_{ij} - \tilde{Y}_{ij} = {} & \{Y_{ij} - \mathsf{E}(Y_{ij})\} - \exp(\omega^2/2)\{\exp(X'_{ij}\hat{\xi}) - \exp(X'_{ij}\xi)\} \\ & + \{\exp(\omega^2/2) - \exp(s^2/2)\}\exp(X'_{ij}\hat{\xi}), \end{aligned} \tag{18}$$

which we will sum over $\mathcal{A}$. This is sometimes called the forecast taxonomy. This expansion gives some insight into the asymptotic forecast distribution, although the detailed proof will be left to the appendix. The first term in (18) is the process error. When extending Theorem 3.4 to the lower triangle $\mathcal{J}$ we will get

$$\omega^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, \varsigma^2_{\mathcal{A},process}), \tag{19}$$

where

$$\varsigma^2_{\mathcal{A},process} = \sum_{(i,j)\in\mathcal{A}} \exp(2X'_{ij}\xi) \tag{20}$$

The second term in (18) is the estimation error for the canonical parameter ξ. From Theorem 3.5 we will be able to derive

$$\omega^{-1}\exp(\omega^2/2)\{\exp(X'_{ij}\hat{\xi}) - \exp(X'_{ij}\xi)\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, \varsigma^2_{\mathcal{A},estimation}), \tag{21}$$

where

$$\varsigma^2_{\mathcal{A},estimation} = \left\{\sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\xi)X'_{ij}\right\}(X'X)^{-1}\left\{\sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\xi)X_{ij}\right\}. \tag{22}$$

The third term in (18) vanishes asymptotically. We will estimate ω^2 by s^2, which turns the asymptotic normal distributions into t-distributions. The process error and the estimation error are asymptotically independent as they are based on independent variables for the upper and lower triangle, $\mathcal{J}$ and $\mathcal{I}$. We can describe the asymptotic forecast error as follows.

Theorem 3.8: *Suppose the generalized log-normal model defined by Assumptions* 3.1 *and* 3.2 *applies both in the upper and the lower triangle, $\mathcal{I}$ and $\mathcal{J}$. Then, as $\omega^2 \to 0$,*

$$\hat{\omega}^{-1}(Y_{\mathcal{A}} - \tilde{Y}_{\mathcal{A}}) \overset{\mathsf{D}}{\to} (\varsigma^2_{\mathcal{A},process} + \varsigma^2_{\mathcal{A},estimation})^{1/2}\mathsf{t}_{n-p},$$

where $\varsigma^2_{\mathcal{A},process}$ and $\varsigma^2_{\mathcal{A},estimation}$ can be estimated consistently by

$$r^2_{\mathcal{A},process} = \sum_{(i,j)\in\mathcal{A}} \exp(2X'_{ij}\hat{\xi}), \tag{23}$$

$$r^2_{\mathcal{A},estimation} = \left\{\sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\hat{\xi})X'_{ij}\right\}(X'X)^{-1}\left\{\sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\hat{\xi})X_{ij}\right\}. \tag{24}$$

Thus, the distribution forecast is

$$\tilde{Y}_{\mathcal{A}} + \{\hat{\omega}^2(r^2_{\mathcal{A},process} + r^2_{\mathcal{A},estimation})\}^{1/2}\mathsf{t}_{n-q}. \tag{25}$$

3.4. Specification test

Specification tests for the log-normal model can be carried out by allowing a richer structure for the predictor or for the variance. We have already seen how the generalized log-normal chain-ladder model can be tested against the extended chain-ladder model using an asymptotic F-test. We can test whether the variance is constant across the upper triangle by adopting the Bartlett (1937) test. Recently, Harnau (2018a) has shown how to do model specification tests for the over-dispersed Poisson model. Here we will adapt the Bartlett test to the log-normal chain-ladder. It should be noted that one can of course also allow a richer structure for the predictor and the variance simultaneously following the principles outlined here.

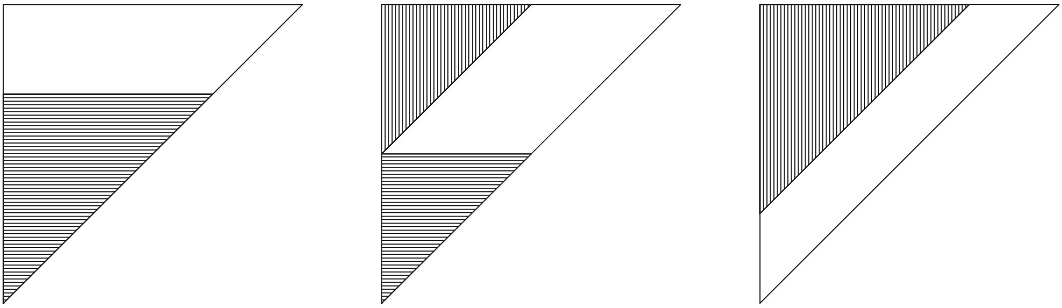

Figure 1. Examples of dividing triangles in two or three parts.

Suppose the triangle $\mathcal{I}$ can be divided into two or more groups as indicated in Figure 1. Thus, the index set $\mathcal{I}$ is divided into disjoint sets $\mathcal{I}_\ell$ for $\ell = 1, \dots, m$. We then set up a log-normal chain-ladder separately for each group. Note that the full canonical parameter vector ξ may not be identified on the subsets. As we will only be interested in the fit of the models we can ad hoc identify ξ by dropping sufficiently many columns of the design matrix X. This gives us a parameter ξ_ℓ and a design vector $X_{ij\ell}$ for each subset $\mathcal{I}_\ell$ and a predictor $\mu_{ij\ell} = X'_{ij\ell}\xi_\ell$. Thus the model for each group is that $y_{ij\ell}$ is $\mathsf{N}(\mu_{ij\ell}, \omega_\ell^2)$. Let p_ℓ denote the dimension of these vectors, while n_ℓ is the number of elements in $\mathcal{I}_\ell$ giving the degrees of freedom $df_\ell = n_\ell - p_\ell$.

When fitting the log-normal chain-ladder separately to each group we get estimators $\hat{\xi}_\ell$ and predictors $\hat{\mu}_{ij\ell} = X'_{ij\ell}\hat{\xi}_\ell$. From this, we can compute the residual sum of squares and variance estimators as

$$RSS_\ell = \sum_{(i,j)\in\mathcal{I}_\ell} (y_{ij} - \hat{\mu}_{ij,\ell})^2, \quad s_\ell^2 = \frac{1}{df_\ell} RSS_\ell. \tag{26}$$

If there are only two subsets then we have two choices of tests available. The first test is a simple F-test for the hypothesis that $\omega_1 = \omega_2$. In the log-normal model this is

$$F^\omega = s_2^2/s_1^2 \overset{\mathsf{D}}{=} \mathsf{F}_{n_2-p_2, n_1-p_1}. \tag{27}$$

In the generalized log-normal the F-distribution can be shown to be valid asymptotically. Harnau (2018a) has proved this for the over-dispersed Poisson model using an infinitely divisible setup. That proof extends to the generalized log-normal setup following the ideas of the proofs of the above theorems. We can then construct a two-sided test. Choosing a 5% level this test rejects when F^ω is either smaller than the 2.5% quantile or larger than the 97.5% quantile of the $\mathsf{F}_{n_2-p_2,n_1-p_1}$ -distribution.

The second test is known as Bartlett's test and applies to any number of groups. Thus, suppose we have m groups and want to test $\omega_1 = \cdots = \omega_m$. In the exact log-normal case then $s_1^2, \dots, s_m^2$ are independent scaled χ^2 variables. Bartlett found the likelihood for this χ^2 model. Under the hypothesis, the common variance is estimated by

$$\bar{s}^2 = \frac{1}{df.} \sum_{\ell=1}^m RSS_\ell, \quad \text{where} \quad df. = \sum_{\ell=1}^m df_\ell = n - \sum_{\ell=1}^m p_\ell, \tag{28}$$

while the likelihood ratio test statistic for the hypothesis is

$$LR^\omega = df. \log(\bar{s}^2) - \sum_{\ell=1}^m df_\ell \log(s_\ell^2). \tag{29}$$

The exact distribution of the likelihood ratio test statistic depends on the degrees of freedom of the groups, but not on their ordering. No analytic expression is known. However, Bartlett showed that this distribution is very well approximated by a scaled χ^2-distribution. That is

$$\frac{LR^{\omega}}{C} \approx \chi^2_{m-1} \quad \text{where} \quad C = 1 + \frac{1}{3(m-1)} \left(\sum_{\ell=1}^{m} \frac{1}{df_\ell} - \frac{1}{df_\cdot} \right). \tag{30}$$

The factor C is known as the Bartlett correction factor. Formally, the approximation is a second order expansion which is valid when the small group is large, so that $\min_\ell df_\ell$ is large. However, the approximation works exceptionally well in very small samples; see the simulations by Harnau (2018a). Once again the Bartlett test (29) will be applicable in the generalized log-normal model, which can be proved by following the proof of Harnau (2018a).

In practice, we can fit separate log-normal models to each group, that is $y_{ij\ell}$ is assumed $\mathsf{N}(\mu_{ij\ell}, \omega^2_\ell)$. If the Bartlett test does not reject the hypothesis of common variance we then arrive at a model where $y_{ij\ell}$ is assumed $\mathsf{N}(\mu_{ij\ell}, \omega^2)$. This model can be estimated by a single regression where the design matrix is block diagonal, $X^m = \text{diag}(X_1, X_2, \ldots, X_m)$ of dimension $p_\cdot = \sum_{\ell=1}^{m} p_\ell$. We then compare the models with design matrices X^m and the original X of the maintained model through an F-test.

4. Empirical illustration

We apply the theory to the insurance run-off triangle shown in Table 1. All R Core Team (2019) code is given in the supplementary material. We use the R packages `apc` version 1.3.5, see Nielsen (2015) and `ChainLadder`, see Gesmann et al. (2015). First, we apply the proposed inference and estimation procedures to the data. This is followed first by distribution forecast and then by an analysis of the model specification.

4.1. Inference and estimation

We apply the log-normal model to the data and consider three nested parametrizations:

apc age-period-cohort model = extended chain-ladder
ac age-cohort model = chain-ladder
ad age-drift model = chain-ladder with a linear accident year effect

Table 2 shows an analysis of variance. This conforms with the exact distribution theory in Section 2.3 and the asymptotic distribution theory in Theorems 3.5 and 3.7 in Section 3.2.

First, we test the chain-ladder model (ac for age-cohort) against the extended chain-ladder model (apc for age-period-cohort) with p -value of $p = .984$. The chain-ladder hypothesis is clearly not rejected at a conventional 5% test level. Next, we test the further restriction (ad for age-drift) that the row differences are constant, that is $\Delta^2\alpha_i = 0$. We get $p = .000$ both when testing against the apc and the ac model. This suggests that a further reduction of the model is not supported. In summary, the analysis of variance indicates that it is adequate to proceed with a chain-ladder specification and thereby ignore calendar effects.

Table 2. Analysis of variance for the US casualty data.

sub	$-2\log L$	df_{sub}	$F_{sub,apc}$	*P*	$F_{sub,ac}$	*P*
apc	170.003	153				
ac	179.873	171	0.41	.984		
ad	258.570	189	2.23	.000	4.32	.000

Table 3. Estimates for the US casualty data for the log-normal chain-ladder (ac).

	Estimate	se_t		Estimate	se_t
μ_{11}	7.660	0.138			
$\Delta\alpha_2$	0.289	0.134	$\Delta\beta_2$	2.272	0.134
$\Delta\alpha_3$	0.163	0.136	$\Delta\beta_3$	0.933	0.136
$\Delta\alpha_4$	−0.265	0.140	$\Delta\beta_4$	0.236	0.140
$\Delta\alpha_5$	0.150	0.144	$\Delta\beta_5$	0.089	0.144
$\Delta\alpha_6$	−0.374	0.148	$\Delta\beta_6$	−0.176	0.148
$\Delta\alpha_7$	−0.199	0.153	$\Delta\beta_7$	−0.144	0.153
$\Delta\alpha_8$	−0.009	0.159	$\Delta\beta_8$	−0.428	0.159
$\Delta\alpha_9$	−0.005	0.165	$\Delta\beta_9$	−0.301	0.165
$\Delta\alpha_{10}$	−0.132	0.172	$\Delta\beta_{10}$	−0.400	0.172
$\Delta\alpha_{11}$	−0.022	0.180	$\Delta\beta_{11}$	−0.190	0.180
$\Delta\alpha_{12}$	−0.473	0.190	$\Delta\beta_{12}$	−0.242	0.190
$\Delta\alpha_{13}$	−0.438	0.200	$\Delta\beta_{13}$	−0.260	0.200
$\Delta\alpha_{14}$	0.296	0.214	$\Delta\beta_{14}$	−0.555	0.214
$\Delta\alpha_{15}$	0.311	0.230	$\Delta\beta_{15}$	−0.303	0.230
$\Delta\alpha_{16}$	−0.269	0.250	$\Delta\beta_{16}$	0.406	0.250
$\Delta\alpha_{17}$	0.142	0.277	$\Delta\beta_{17}$	−0.895	0.277
$\Delta\alpha_{18}$	0.202	0.316	$\Delta\beta_{18}$	0.117	0.316
$\Delta\alpha_{19}$	−0.093	0.378	$\Delta\beta_{19}$	−0.383	0.378
$\Delta\alpha_{20}$	0.873	0.508	$\Delta\beta_{20}$	−0.273	0.508
s^2	0.169		*RSS*	28.956	

Table 3 shows the estimated parameters for the log-normal model with chain-ladder structure (ac). We report standard errors se_t following Theorem 3.6. They are the same for $\Delta\alpha$ and $\Delta\beta$ due to symmetry of $(X'X)^{-1}$ at the diagonal. These follow a t-distribution with $n-p=171$ degrees of freedom, since the triangle has dimension $k=20$ and $n = k(k+1)/2 = 210$ and $p=2k-1=39$. The corresponding two-sided 95% critical values are 1.97. We also report the degrees of freedom corrected estimate, s^2, for ω^2. We see that many of the development year effects $\Delta\beta$, in particular $\Delta\beta_2$, are significant. The first few development year effects are positive, which matches the increases seen in first few columns of the data in Table 1. At the same time, many accident year effects $\Delta\alpha$ are not individually significant, although they are jointly significant as seen in Table 2. The signs of the $\Delta\alpha$'s match the relative increase or decrease of the amounts seen in the rows of Table 1.

In Appendix 2, we present a further Table A1 with estimates. These are the estimated parameters for the log-normal model with an extended chain-ladder structure (apc) as in Section 2.5. These will be used for the simulation study. The $\Delta^2\gamma$-coefficients measure the calendar effect and are restricted to zero in the chain-ladder model.

4.2. Distribution forecasting

Table 4 shows forecasts of reserves for the US casualty data in different accident years, i.e. the row sums in the lower triangle $\mathcal{J}$. We report results from the generalized log-normal chain-ladder model (GLN), the over-dispersed Poisson chain-ladder (ODP) and England (2002) bootstrap (BS). For each method, we present a point forecast of the reserve, the standard error over point forecast (*se*/Res) and the 1 in 200 over point forecast values (99.5%/Res).

For the generalized log-normal chain-ladder model, we use the asymptotic distribution forecast in (25). For the over-dispersed Poisson model, we use the asymptotic distribution forecasts from Harnau & Nielsen (2018, Equation (11)). For the bootstrap, we use the ChainLadder package by Gesmann et al. (2015), based on the method described in England (2002). We apply 10^5 bootstrap draws using the gamma option. We also applied the ODP option and got similar results, which are therefore not reported here.

Table 4 shows that the over-dispersed Poisson forecasts are similar to the bootstrap. Their point forecasts are smaller than that of the generalized log-normal model. This is in part due to the

Table 4. Forecasting for the US casualty data using the generalized log-normal, the over-dispersed Poisson model and the bootstrap.

	Generalized log-normal			Over-dispersed Poisson			Bootstrap		
i	Reserve	$\frac{se}{\text{Res}}$	$\frac{99.5\%}{\text{Res}}$	Reserve	$\frac{se}{\text{Res}}$	$\frac{99.5\%}{\text{Res}}$	Reserve	$\frac{se}{\text{Res}}$	$\frac{99.5\%}{\text{Res}}$
2	1871	0.55	2.43	1368	1.81	5.71	1345	1.99	9.93
3	5099	0.37	1.96	4476	0.92	3.40	4415	0.97	4.63
4	7171	0.30	1.77	6925	0.69	2.78	6830	0.71	3.56
5	11,699	0.26	1.66	10,975	0.54	2.41	10,846	0.56	2.90
6	13,717	0.24	1.64	14,941	0.44	2.14	14,767	0.45	2.50
7	14,344	0.22	1.58	18,337	0.39	2.01	18,147	0.40	2.29
8	18,377	0.21	1.54	24,487	0.34	1.87	24,233	0.35	2.09
9	25,488	0.21	1.54	31876	0.29	1.76	31607	0.30	1.93
10	30,525	0.20	1.53	35,567	0.28	1.72	35,270	0.28	1.87
11	40,078	0.20	1.53	48,595	0.24	1.63	48,176	0.25	1.73
12	32,680	0.20	1.53	42,027	0.26	1.68	41,659	0.27	1.80
13	28,509	0.21	1.54	37,114	0.28	1.74	36,814	0.29	1.88
14	51,761	0.21	1.55	66,977	0.22	1.58	66,554	0.23	1.69
15	98,748	0.22	1.58	102982	0.20	1.51	102282	0.20	1.59
16	100,331	0.23	1.60	136647	0.19	1.51	135880	0.20	1.59
17	149,813	0.24	1.64	164,318	0.22	1.56	163,500	0.22	1.68
18	221,550	0.26	1.69	218,874	0.25	1.66	218,115	0.26	1.83
19	229,481	0.30	1.79	166,120	0.49	2.29	166,431	0.51	2.84
20	575,343	0.41	2.06	337,001	0.94	3.46	353,628	1.03	4.91
Total	1,656,586	0.16	1.42	1,469,605	0.23	1.60	1,480,500	0.26	1.95

Note: The bootstrap simulation is based on 10^5 repetitions.

additional factor $\exp(s^2/2) = \exp(0.169/2) = 1.088$ in the generalized log-normal point forecast. The difference seems large compared to the authors' experience with other data. It is possibly due to the relatively large dimension of the triangle, so that there are more degrees of freedom to pick up differences between the over-dispersed Poisson and the generalized log-normal models.

The standard error and 99.5% quantiles over reserve ratios are generally lower and less variable for the generalized log-normal chain-ladder model. This is especially pronounced for early accident years and the latest accident year.

Figure 2 shows the trends of the reserve and standard error and 99.5% quantile over reserve ratios for the three methods. The point forecast trends are similar for models, showing an increasing trend with accident year as expected. The ratios are seen to be flatter for the generalized log-normal model.

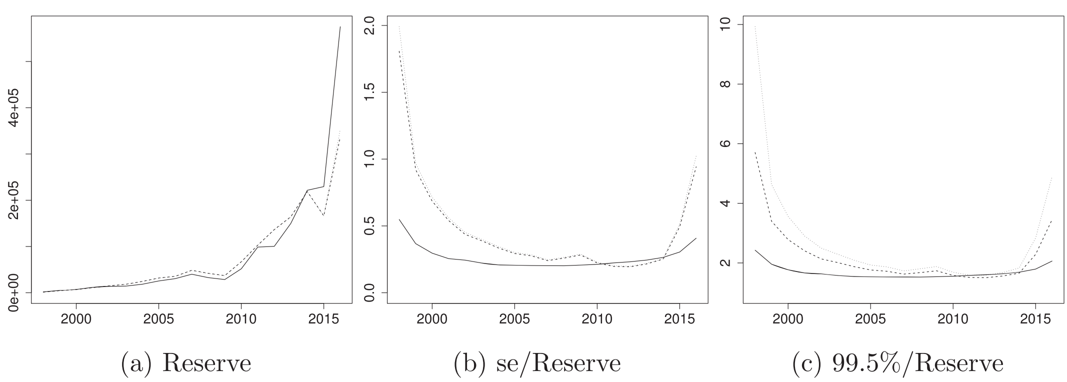

Figure 2. Illustration of the forecasts in Table 4 for the US casualty data. Solid line is the generalized log-normal forecast. Dashed line is the over-dispersed Poisson forecast. Dotted line is the bootstrap forecast. Panel (a) shows the reserves against accident year i. Panel (b) shows the standard error to reserve ratio. Panel (c) shows the 99.5% quantile to reserve ratio. (a) Reserve. (b) se/Reserve. (c) 99.5%/Reserve.

This is related to the assumption of the generalized log-normal chain-ladder model that standard deviation to mean ratio is constant across the entries, while the variance to mean ratio is assumed constant for the over-dispersed Poisson model and the bootstrap.

4.3. Recursive distribution forecasting

To check the robustness of the model, we apply the distribution forecasting recursively. Thus, we apply the distribution forecast to subsets of the triangle.

In this way, Table 5 shows standard error and 99.5% over reserve ratios. It has nine panels, where the rows are for the asymptotic generalized log-normal model, the over-dispersed Poisson model and the bootstrap, respectively. In the first column, we show the ratios for the last five accident years based on the full triangle. These numbers are the same as those in Table 4. In the second column, we omit the last diagonal of the data triangle to get a $k-1=19$ dimensional triangle. We then forecast the last five accident years relative to that triangle. In the third column, we omit the last two diagonals of the data triangle to get a $k-2=18$ dimensional triangle.

We see that the generalized log-normal forecasts are stable for all years. The over-dispersed Poisson and bootstrap forecasts are less stable in the latest accident year. This is possibly because of instability in the corners of the data triangle shown in Table 1, that may be dampened when taking logs. Alternatively, it could be attributed to a better fit of the log-normal model across the entire triangle. We will explore the model specification using formal tests in the next section.

4.4. Model selection

We now apply the specification test outlined in Section 3.4 for the log-normal model and in Harnau (2018a) for the over-dispersed Poisson model. For the tests, we split the data triangle of Table 1 as outlined in Figure 1:

Table 5. Recursive forecasting for the US casualty data in the latest five accident years.

Full triangle			Leave 1 out			Leave 2 out		
i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$	i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$	i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$
Generalized log-normal								
16	0.23	1.60	15	0.23	1.61	14	0.23	1.61
17	0.24	1.64	16	0.25	1.64	15	0.25	1.64
18	0.26	1.69	17	0.27	1.69	16	0.27	1.69
19	0.30	1.79	18	0.31	1.80	17	0.31	1.80
20	0.41	2.06	19	0.41	2.07	18	0.41	2.07
all	0.16	1.42	all	0.13	1.33	all	0.12	1.31
over-dispersed Poisson								
16	0.19	1.51	15	0.20	1.53	14	0.22	1.58
17	0.22	1.56	16	0.22	1.56	15	0.24	1.62
18	0.25	1.66	17	0.28	1.74	16	0.28	1.72
19	0.49	2.29	18	0.48	2.25	17	0.48	2.24
20	0.94	3.46	19	1.38	4.61	18	1.51	4.94
all	0.23	1.60	all	0.20	1.53	all	0.20	1.52
bootstrap								
16	0.20	1.59	15	0.21	1.62	14	0.23	1.70
17	0.22	1.68	16	0.22	1.68	15	0.24	1.75
18	0.26	1.83	17	0.29	1.97	16	0.28	1.92
19	0.51	2.84	18	0.49	2.78	17	0.49	2.77
20	1.03	4.91	19	1.49	6.69	18	1.66	7.45
all	0.26	1.95	all	0.23	1.81	all	0.22	1.79

Note: The bootstrap simulation is based on 10^5 repetitions.

Table 6. Bartlett tests for common dispersion and F tests for common mean parameters.

	Generalized log-normal				Over-dispersed Poisson			
Splits	LR^{ω}/C	P	F	P	LR^{ω}/C	P	F	P
(a)	6.29	0.012	5.50	.000	11.68	.001	6.62	.000
(b)	4.70	0.095	4.48	.000	11.63	.003	6.03	.000
(c)	1.12	0.291	3.08	.000	15.07	.000	2.50	.000

(a) a horizontal split with the first 6 rows in one group and the last 14 rows in a second group.
(b) a horizontal and diagonal split with the first 10 diagonals in one group, the last 10 rows in a second group and the remaining entries in a third group.
(c) a diagonal split with the first 14 diagonals in one group and the last 6 diagonals in a second group.

For each split, we estimate a chain-ladder structure separately for each subgroup. Table 6 reports the Bartlett test statistic LR^{ω}/C from (30) for a common variance across groups. Given a common variance we also compute an F-statistic for common chain-ladder structure in the mean.

For each of the generalized log-normal and over-dispersed Poisson model, we are conducting six tests. When chosing the size of each individual test, that is the probability of falsely rejecting the hypothesis, we would have to keep in mind the overall size of rejecting any of the hypotheses. If the test statistics were independent and the individual tests were conducted at level p the overall size would be $1-(1-p)^6 \approx 6p$ by binomial expansion, see also Hendry & Nielsen (2007, Section 9.5). Thus, if the individual tests are conducted at a 1% level we would expect the overall size to be about 5%. At present we have no theory for a more formal calculation of the joint size of the tests.

Starting with the Bartlett test for constant variance, we see that there is little evidence against the log-normal model (columns 1, 2), whereas the over-dispersed Poisson model is rejected (columns 5, 6). The F-tests (columns 3, 4, 7, 8) for change in the predictor μ_{ij} indicate problems in both cases. These F-tests compare models with different parameters in different parts of triangle to the standard Chain-Ladder parameterization. In particular, those models will have more development (age) parameters. This is perhaps not surprising given the long-time span. However, as we saw in Table 2, this is not captured by a calendar effect. It is up to users to choose whether to include the whole 20 years experience or just the recent years. However, in a long tail business like Casualty, one could be better off to include longer history; while if there has been a legislation change it is better to put more weight on the recent years' experience. Going forward, it would be useful to develop more flexible models. Allowing different development parameters in the top and the bottom part of the triangle would certainly be possible within the present framework. If both models had been supported they could be compared using the encompassing test of Harnau (2018b).

5. Simulation

In Theorems 3.7 and 3.8, we presented asymptotic results for inference and distribution forecasting. We now apply simulation to investigate the quality of these asymptotic approximations.

5.1. Test statistic

We assess the finite sample performance of the F-tests proposed in Theorem 3.7 and applied in Table 2. We simulate under the null hypothesis of a chain-ladder specification, ac, as well as under the alternative hypothesis of an extended chain-ladder specification, apc. We choose the distribution to be log-normal so, to be specific, we actually illustrate the well-known exact distribution theory for regression analysis. Theorem 3.7 also applies for infinitely divisible distributions that are not log-normal but satisfy Assumptions 3.1 and 3.2. Such infinitely divisible distributions are, however, not easily generated. The real point of the simulations is therefore to illustrate the small variance asymptotics in Theorem 3.7 by showing that power increases with shrinking variance.

The data generating processes are constructed from the US casualty data as follows. We consider a $k = 20$ dimensional triangle. We assume that the variables Y_{ij} in the upper triangle $\mathcal{I}$ are independent log-normal distributed, so that $y_{ij} = \log(Y_{ij})$ is normal with mean μ_{ij} and variance σ^2. Under the null hypothesis of a chain-ladder specification, H_{ac}, then μ_{ij} is defined from (5) where the parameters μ_{ij} are chosen to match those of Table 3. We also choose σ^2 to match the estimate s^2 from Table 3, but multiplied by a factor v^2 where v is chosen as $2, 1, 1/2$ to capture the small-variance asymptotics. Under the alternative, we apply the extended chain-ladder specification H_{apc} where the parameters are chosen to match those of Table A1. In all cases, we draw 10^5 repetitions.

We note that the $F_{(18,153)}$ -distribution is exact under the null hypothesis, since we are operating on the log-scale and simulate normal variables so that standard regression theory applies. Indeed, Table 7 shows that simulated size (type I error) is correct apart from Monte Carlo standard error. We check this for at the 1%, 5% and 10% level for $v = 2, 1, 1/2$.

Under the alternative we simulate power (unity minus type II error). The exact distribution is a non-central F -distribution. The simulations show that the power increases for shrinking variance $v^2\omega^2$ and for increasing level (type I error) of the test.

We can also illustrate the increasing power with shrinking variance through the following analytic example. Suppose we consider variables $Z_1, \ldots, Z_n$ that are independent $\mathsf{N}(\mu, \omega^2)$-distributed. Then the parameters are estimated by $\hat{\mu} = \bar{Z}$ and $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$. The t-statistic for $\mu = 0$ has the expansion

$$\frac{\hat{\mu} - 0}{\sqrt{s^2/(n-1)}} = \frac{\hat{\mu} - \mu}{\sqrt{s^2/(n-1)}} + \frac{\mu - 0}{\sqrt{s^2/(n-1)}}.$$

The first term is t distributed with $(n-1)$ degrees of freedom regardless of the value of μ. The second term is zero under the hypothesis $\mu = 0$. Under the alternative $\mu \neq 0$ the second term is non-zero and measures non-centrality so that the overall t -statistic is non-central t. In standard asymptotic theory n is large so that for fixed μ, ω then s^2 is consistent for ω^2 and the second term is close to $\mu/\sqrt{\omega^2/(n-1)} = (\mu/\omega)\sqrt{(n-1)}$. Due to the $(n-1)$-factor the non-centrality diverges, so that the power increases to unity and the test is consistent. In the small variance asymptotics ω^2 shrinks to zero while n is fixed. Then s^2 vanishes, see Theorem 3.7, and the non-centrality diverges in a similar way even though n is fixed.

In practice, quarterly data may be available. Thus, suppose the sample period otherwise is fixed and the triangle has four times as many accident and development periods. This means that the available information will spread over up to 16 as many cells so that σ^2 would increase. The second effect change was investigated in Table 7, indicating that the results would become more noisy in that there would be a deterioration in the power to reject an incorrect null hypothesis of no calendar effect. To investigate the first effect we ran the experiment of Table 7 for different data generating processes. The triangle above has dimension 20, that is 18 double differences in each dimension. We considered triangles of dimension 11 and 37, that is with half and with twice as many double differences. These double differences were chosen by first computing the double sums of the double differences and then

Table 7. Simulated performance of F test based on 10^5 draws.

	Size under H_{ac}			Power under H_{apc}		
Confidence level	1.00%	5.00%	10.00%	1.00%	5.00%	10.00%
$v = 2$	1.01%	5.00%	10.16%	2.26%	9.03%	16.31%
$v = 1$	0.98%	5.07%	10.07%	10.49%	27.51%	40.22%
$v = 0.5$	0.99%	5.09%	10.05%	78.03%	92.17%	96.07%

Note: The Monte Carlo standard error is less than 0.1 %.

Table 8. Simulated power of F test under H_{apc} based on 10^5 draws.

	Dimesion 11			Dimension 37		
Confidence level	1.00%	5.00%	10.00%	1.00%	5.00%	10.00%
$v = 2$	1.45%	6.65%	12.61%	3.69%	13.05%	21.89%
$v = 1$	3.24%	12.14%	20.91%	29.58%	54.11%	67.14%
$v = 0.5$	19.62%	43.26%	57.60%	99.71%	99.97%	99.99%

Note: The Monte Carlo standard error is less than 0.1 %.

either taking every other value to get 9 values or linearly interpolate to get 35 values. This would then be double differenced again. The results are reported in Table 8. The F -distribution is exact as before. We therefore only report the power. It appears that for fixed dispersion $v\sigma$ the power is increasing with the dimension so that the previous effect is offset to some extent. Overall, we would expect poorer performance of distribution approximations with quarterly data than with annual data. This could possibly be offset if there were seasonal patterns that could be exploited.

5.2. Forecasting

We assess the finite sample performance of the asymptotic distribution forecasts proposed in Theorem 3.8 and applied in Table 4. These asymptotic distribution forecasts are compared to the over-dispersed Poisson forecast of Harnau & Nielsen (2018) and the bootstrap of England & Verrall (1999) and England (2002). Two different log-normal chain-ladder data generating processes are used. First, we apply the estimates from the US casualty data so that the parameters are chosen to match those of Table 3. As before the variance ω^2 is multiplied by a factor v^2 where $v = 2, 1, 1/2$. We have seen that the over-dispersed Poisson model is poor for this data set and we will expect the generalized log-normal distribution forecasts to be superior. Secondly, we obtain similar estimates for the Taylor & Ashe (1983) data, see also Harnau & Nielsen (2018, Table 1). For those data the generalized log-normal model and the over-dispersed Poisson model provide equally good fits so that the different distributions forecasts should be more similar in performance.

We first compare the asymptotic distribution forecast from Theorem 3.8 with the exact forecast distribution. This is done by simulating log-normal chain-ladder for both the upper and the lower triangles, $\mathcal{I}$ and $\mathcal{J}$. The true forecast error distribution is then based on $Y_\mathcal{A} - \tilde{Y}_\mathcal{A}$, where $Y_\mathcal{A}$ is computed from the simulated lower triangle $\mathcal{J}$ while $\tilde{Y}_\mathcal{A}$ is the log-normal point forecast computed from the upper triangle data $\mathcal{I}$. We compute the true forecast error $Y_\mathcal{A} - \tilde{Y}_\mathcal{A}$ for each simulation draw and report mean, standard error and quantiles of the draws. This is done for the entire reserve, so that $\mathcal{A} = \mathcal{J}$. The asymptotic theory in Theorem 3.8 provides a t -approximation, so that for each draw of the upper triangle $\mathcal{I}$, we also compute mean, standard error and quantiles from the t -approximations and report averages over the draws.

The first panel of Table 9 compares the simulated actual forecast distribution, $true^{GLN}$, with the simulated t -approximations, t^{GLN}. We see that with shrinking variance factor v then the overall forecast distribution becomes less variable and the t -approximation becomes relatively better. The t -approximation is symmetric and does not fully capture the asymmetry of the actual distribution. We note that the performance of the t -approximation is better in the upper tail than the lower tail, which is beneficial when we are interested in 99.5% value at risk.

The second panel of Table 9 shows the performance of the traditional chain-ladder. Since the data are log-normal we expect the chain-ladder to perform poorly. We apply the asymptotic theory of Harnau & Nielsen (2018) and the bootstrap of England & Verrall (1999) and England (2002) as implemented by Gesmann et al. (2015). The results are generated as before with the difference that the point forecasts are based on the traditional chain-ladder, while the data remain log-normal. The actual forecast errors, $true^{ODP}$ are similar to the previous actual errors $true^{GLN}$, particular in the right

Table 9. Simulation performance of distribution forecasts for the US casualty data.

v		Moments		Quantiles						
		Mean	SE	0.5%	1%	5%	50%	95%	99%	99.5%
	generalized log-normal (GLN)									
2	$true^{GLN}$	3.0	12.6	−55.1	−42.6	−18.5	5.4	17.2	22.2	24.4
	t^{GLN}	0.0	7.9	−20.7	−18.7	−13.1	0.0	13.1	18.7	20.7
1	$true^{GLN}$	0.5	3.3	−11.2	−9.5	−5.5	0.9	5.0	6.5	7.0
	t^{GLN}	0.0	3.0	−7.7	−6.9	−4.9	0.0	4.9	6.9	7.7
0.5	$true^{GLN}$	0.1	1.4	−4.1	−3.6	−2.3	0.2	2.3	3.0	3.3
	t^{GLN}	0.0	1.4	−3.6	−3.2	−2.3	0.0	2.3	3.2	3.6
	over-dispersed Poisson (ODP) and bootstrap (BS)									
2	$true^{ODP}$	7.7	10.5	−37.9	−28.5	−10.0	9.3	20.3	25.4	27.3
	t^{ODP}	0.0	19.8	−51.6	−46.5	−32.8	0.0	32.8	46.5	51.6
	BS	−15.4	2631.6	−683.1	−350.8	−78.9	3.3	55.8	313.3	643.1
1	$true^{ODP}$	1.3	3.2	−9.9	−8.3	−4.5	1.7	5.8	7.3	7.8
	t^{ODP}	0.0	7.9	−20.7	−18.6	−13.1	0.0	13.1	18.6	20.7
	BS	−1.8	123.4	−73.9	−50.1	−21.2	0.5	12.5	23.4	35.1
0.5	$true^{ODP}$	0.3	1.4	−4.0	−3.5	−2.2	0.4	2.5	3.3	3.6
	t^{ODP}	0.0	3.8	−9.8	−8.8	−6.2	0.0	6.2	8.8	9.8
	BS	−0.2	4.2	−15.4	−13.1	−7.5	0.1	5.9	9.1	10.3
	root-mean-square-errors (rms)									
2	rms^{GLN}	3.0	8.3	38.7	28.8	12.5	5.4	11.9	16.3	18.1
	rms^{ODP}	7.7	13.8	29.7	29.9	28.2	9.3	20.9	31.8	35.9
	rms^{BS}	4284.4	135397.1	925.7	431.1	86.4	6.8	17.3	52.7	397.7
1	rms^{GLN}	0.5	1.1	4.5	3.6	1.9	0.9	1.8	2.6	2.9
	rms^{ODP}	1.3	5.1	11.9	11.3	9.2	1.7	8.0	12.2	13.8
	rms^{BS}	67.6	2132.3	79.5	48.4	18.2	1.2	5.4	6.1	18.8
0.5	rms^{GLN}	0.1	0.3	0.8	0.7	0.4	0.2	0.4	0.6	0.7
	rms^{ODP}	0.3	2.4	5.9	5.5	4.1	0.4	3.8	5.7	6.4
	rms^{BS}	0.6	3.0	11.9	10.0	5.5	0.3	2.4	2.7	5.7

Notes: Results in USD. The study is based on 10^5 repetitions, and for each simulated upper triangle, the bootstrap is based on 999 simulations.

tail of the distribution. The asymptotic distribution approximation, t^{ODP}, and the bootstrap approximation, *BS*, do not provide the same quality of approximations as t^{GLN} did for $true^{GLN}$. For large $v = 2$, the bootstrap is very poor, possibly because of resampling of large residuals arising from the mis-specification.

We also simulate the root mean square forecast error for the three methods. For the log-normal asymptotic distribution approximation this is computed as follows. We first find mean, standard deviation and quantiles of the infeasible reserve based on the draws of the lower triangle $\mathcal{J}$. This is the true forecast distribution. For each draw of the upper triangle $\mathcal{I}$ we then compute mean, standard deviation and quantiles of the asymptotic distribution forecast (25) and subtract the mean, standard deviation and quantiles, respectively, of the true forecast distribution. We square, take average across the draws of the upper triangle $\mathcal{I}$, and then the take the square root. Similar calculations are done for the over-dispersed approximation and the bootstrap.

The third panel of Table 9 shows the root mean square forecast errors. We see that the generalized log-normal distribution approximation is superior in all cases and that the bootstrap can be very poor if v is not small.

In Table 10, we repeat the simulation exercise for the Taylor & Ashe (1983) data. For these data, we repeated the empirical exercise of Section 4, although we do not report the results here. We found that the generalized log-normal chain-ladder and the over-dispersed chain-ladder appear to give equally good fit, so that we will expect less difference between the methods in this case. We suspect that this arises because of two features in the data. The Taylor and Ashe triangle has a smaller dimension of $k = 10$ and there is less difference between the accident year parameters, see also Harnau &

Table 10. Simulation performance of distribution forecasts for the data used in Taylor & Ashe (1983).

		Moments		Quantiles						
v		Mean	SE	0.5%	1%	5%	50%	95%	99%	99.5%
	generalized log-normal (GLN)									
2	$true^{GLN}$	7.2	99.8	−372.9	−310.0	−170.0	20.4	140.6	187.5	206.2
	t^{GLN}	0.0	75.7	−205.7	−184.2	−127.7	0.0	127.7	184.2	205.7
1	$true^{GLN}$	1.7	31.8	−96.4	−83.7	−54.0	3.9	49.6	66.8	72.8
	t^{GLN}	0.0	29.7	−80.7	−72.2	−50.1	0.0	50.1	72.2	80.7
0.5	$true^{GLN}$	0.4	14.3	−39.6	−35.4	−23.9	0.9	23.0	31.7	34.4
	t^{GLN}	0.0	14.0	−38.0	−34.0	−23.6	0.0	23.6	34.0	38.0
	over-dispersed Poisson (ODP) and bootstrap (BS)									
2	$true^{ODP}$	45.1	91.4	−297.9	−242.1	−116.8	56.9	168.2	213.5	230.8
	t^{ODP}	0.0	76.6	−208.4	−186.6	−129.4	0.0	129.4	186.6	208.4
	BS	−14.1	340.9	−414.3	−335.9	−193.8	−0.3	114.3	155.6	177.4
1	$true^{ODP}$	9.1	31.9	−89.8	−76.9	−46.8	11.4	56.9	73.5	79.6
	t^{ODP}	0.0	31.7	−86.1	−77.1	−53.5	0.0	53.5	77.1	86.1
	BS	−2.5	35.4	−109.5	−97.2	−64.6	0.1	50.5	68.2	74.1
0.5	$true^{ODP}$	2.1	14.7	−39.3	−34.7	−22.8	2.7	25.2	33.8	36.9
	t^{ODP}	0.0	15.1	−41.2	−36.9	−25.6	0.0	25.6	36.9	41.2
	BS	−0.6	16.5	−46.3	−41.6	−28.6	0.0	25.3	34.9	38.2
	root-mean-square-errors (rms)									
2	rms^{GLN}	7.2	45.3	197.1	156.7	77.4	20.4	66.0	93.4	104.3
	rms^{ODP}	45.1	32.2	118.5	89.1	49.9	56.9	61.9	74.7	80.9
	rms^{BS}	645.6	20322.2	415.1	259.0	126.9	57.4	168.6	107.6	107.7
1	rms^{GLN}	1.7	7.4	24.8	20.7	12.6	3.9	12.0	18.1	20.8
	rms^{ODP}	9.1	6.4	17.9	15.6	12.7	11.4	11.4	16.0	18.7
	rms^{BS}	11.7	8.6	36.0	32.7	23.7	11.3	56.8	25.2	17.7
0.5	rms^{GLN}	0.4	2.2	6.0	5.4	3.6	0.9	3.7	5.7	6.8
	rms^{ODP}	2.1	2.3	6.4	5.9	4.7	2.7	3.8	6.3	7.5
	rms^{BS}	2.7	3.1	11.3	10.3	7.6	2.7	25.1	9.3	5.7

Notes: The study is based on 10^5 repetitions, and for each simulated upper triangle, the bootstrap is based on 999 simulations.

Nielsen (2018, Table 2). As before we simulate a log-normal distribution with parameters equal to the estimates from the data.

Table 10 shows that the three methods perform similarly. In this discussion, we focus on the root mean square error for the 99.5% quantile which is perhaps of most practical interest. For large $v=2$ and $v=1$, the over-dispersed Poisson method actually dominates the generalized log-normal model even though the data are generated to be log-normal. For a smaller $v = 1/2$, the asymptotic approximation for the generalized log-normal beats that of the over-dispersed model slightly. However, the bootstrap appears to be best for $v=1$ and $v = 1/2$.

6. Conclusion

We have presented a new method for distribution forecasting of general insurance reserves in terms of the generalized log-normal model. The forecasts are done under the asymptotic framework which allows users to draw inferences and make model selections easily. This gives an alternative to the traditional chain-ladder where we have the commonly used bootstrap method developed by England & Verrall (1999) and England (2002) along with the recent asymptotic theory of Harnau & Nielsen (2018).

The actuary will have to choose whether the traditional or the log-normal chain ladder or a third method should be used for a given reserving triangle. In some situations, the log-normal chain ladder will be better than the traditional chain latter as shown in our empirical data analysis and simulation study. In addition, we have considered a number of London market datasets. We compared the standard error over mean forecast trends by year of account with the actuaries' selected volatilities and found that the generalized log-normal trends are more in line with the actuaries selected trends than the over-dispersed Poisson model.

The generalized log-normal model distribution forecasts developed here could also improve the actuarial process for a corporation. The log-normal is also often used in simulating attritional reserve risk for capital modelling. At present, this is sometimes combined with the bootstrap method for the traditional chain ladder. This can result in inconsistencies often between reserving and capital modelling.

A limitation of the log-normal model is that it only fits positive incremental values, while in real life some values can be negative due to reinsurance recoveries, salvage or other data issues such as mis-allocation between classes of business or currencies. In these cases, judgements are required and further research must look at how to provide statistical tools to overcome such a limitation.

There is also scope to develop a more advanced model selection process than the model specification tests discussed here. This will give actuaries a statistical basis to select one model over another rather than just eye-balling a distribution fit on a graph. Testing constancy of the dispersion as presented here for the log-normal chain ladder and by Harnau (2018a) for the traditional chain ladder is a beginning of that research agenda.

Chain ladder models involve many parameters, so that the number of parameters is proportional to the dimension of the triangle. A key feature of the asymptotic theory is that the dimension of the triangle is fixed, while the asymptotics is developed along the level of aggregation in the triangle. For a correctly specified model, this, in turn, gives smaller variation, so that the small σ asymptotics applies. If data are disaggregated by quarter or by month we would expect larger triangles. This in itself would not pose any particular computational challenge. However, we would expect more sampling variation, so that σ would be larger and we would expect the small σ asymptotics would tend to work less well. It would be interesting to investigate this further. In particular, if seasonal patterns are common from year to year or accross triangles there may be scope for improving the performance of the asymptotic theory.

The bootstrap method has become popular in recent decades. This is because it usually produces distributions that appear reasonable and it is a simulation based technique which is favoured by many actuaries. A deeper understanding of the bootstrap method can be developed so that it allows model selections and extensions to generate reserve forecasts under other distributions than the over-dispersed Poisson.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding was received from H2020 European Research Council (ERC) grant 694262, DisCont (Nielsen).

References

Barnett, G. & Zehnwirth, B. (2000). Best estimates for reserves. *Proceedings of the Casualty Actuarial Society* **87**, 245–321.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A* **160**, 268–282.

Beard, R. E., Pentikäinen, T. & Pesonen, E. (1984). *Risk theory*, 3rd ed. London: Chapman & Hall.

Doray, L. G. (1996). UMVUE of the IBNR reserve in a lognormal linear regression model. *Insurance: Mathematics and Economics* **18**, 43–57.

Elpidorou, V., Margraf, C., Martínez-Miranda, M. D. & Nielsen, B. (2019). A likelihood approach to Bornhuetter-Ferguson analysis. *Risks* **7**, 119. DOI:10.3390/risks7040119.

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

England, P. & Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics* **25**, 281–293.

England, P. D. & Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal* **8**, 519–44.
Gesmann, M., Murphy, D., Zhang, Y., Carrato, A., Crupi, G., Wüthrich, M. & Concina, F. (2015). Chainladder: statistical methods and models for claims reserving in general insurance. cran.R-project.org/package = ChainLadder.
Harnau, J. (2018a). Misspecification tests for chain-ladder models. *Risk* **6**, 25.
Harnau, J. (2018b). Log-normal over over-dispersed Poisson. *Risk* **6**, 70.
Harnau, J. & Nielsen, B. (2018). Over-dispersed age-period-cohort models. *Journal of the American Statistical Association* **113**, 1722–1732.
Hendry, D. F. & Nielsen, B. (2007). *Econometric modeling*. Princeton, NJ: Princeton University Press.
Hertig, J. (1985). A statistical approach to IBNR-reserves in marine reinsurance. *ASTIN Bulletin* **15**, 171–183.
Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*, Vol. **1**, 2nd ed. New York: Wiley.
Kremer, E. (1982). IBNR-Claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal* **1982**, 47–55.
Kuang, D., Nielsen, B. & Nielsen, J. P. (2008a). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 979–986.
Kuang, D., Nielsen, B. & Nielsen, J. P. (2008b). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 987–991.
Kuang, D., Nielsen, B. & Nielsen, J. P. (2009). Chain-ladder as maximum likelihood revisited. *Annals of Actuarial Science* **4**, 105–121.
Kuang, D., Nielsen, B. & Nielsen, J. P. (2011). Forecasting in an extended chain-ladder-type model. *Journal of Risk and Insurance* **78**, 345–359.
Kuang, D., Nielsen, B. & Nielsen, J. P. (2015). The geometric chain-ladder. *Scandinavian Actuarial Journal* **2015**, 278–300.
Mack, T. (1999). The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. *ASTIN Bulletin* **29**, 361–366.
Martínez-Miranda, M. D., Nielsen, B. & Nielsen, J. P. (2015). Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society Series A* **178**, 29–55.
Nielsen, B. (2015). apc: an R package for age-period-cohort analysis. *R Journal* **7**, 52–64.
R Core Team (2019). R: A language and environment for statistical computing. www.R-project.org.
Renshaw, A. E. (1989). Chain ladder and interactive modelling (Claims reserving and GLIM). *Journal of the Institute of Actuaries* **116**, 559–587.
Taylor, G. C. & Ashe, F. R. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics* **23**, 37–61.
Thorin, O. (1977). On the infinite divisibility of the lognormal distribution. *Scandinavian Actuarial Journal* **1977**, 121–148.
van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.
Verrall, R. J. (1991). On the estimation of reserves from log-linear models. *Insurance: Mathematics and Economics* **10**, 75–80.
Verrall, R. J. (1994). Statistical methods for the chain-ladder technique. *Casualty Actuarial Society Forum* **1994**, 393–446.
XL Group (2017). 2016 global loss triangles. http://phx.corporate-ir.net/phoenix.zhtml?c = 73041\&p = irol-financial reports.
Zehnwirth, B. (1994). Probabilistic development factors with applications to loss reserve variability, prediction intervals, and risk based capital. *Casualty Actuarial Society Forum* **1994**, 447–605.

Appendices

Appendix 1. Proofs of Theorems

Proof of Theorem 3.2: We check Assumption 3.1. Recall that a log-normally distributed variable Y_{ij} is positive, hence non-negative. It is infinitely divisible as shown by Thorin (1977), so that part (i) holds. The array $\mathcal{I}$ is fixed, so that part (ii) holds. For part (iii) note that the skewness is

$$\mathsf{skew}(Y_{ij}) = \{\exp(\omega^2) - 1\}^{1/2}\{\exp(\omega^2) + 2\},$$

see Johnson et al. (1994, Equation (14.9a)). For $\omega \to 0$ we get that $\exp(\omega^2) \to 1$ so that $\mathsf{skew}(Y_{ij}) \to 0$. Hence part (iii) holds.

We check Assumption 3.2. Note that

$$\mathsf{E}(Y_{ij}) = \exp(\mu_{ij} + \omega^2/2),$$
$$\mathsf{Var}(Y_{ij}) = \exp(2\mu_{ij} + \omega^2)\{\exp(\omega^2) - 1\},$$

see Johnson et al. (1994, Equations (14.8a) and (14.8b)). Hence, $\log \mathsf{E}(Y_{ij}) = \mu_{ij} + \omega^2/2$, so that part (i) holds. We have $\omega^2 \to 0$ while ξ is fixed, so that part (ii) holds. Finally, by Taylor expansion for small ω,

$$\frac{\mathsf{Var}(Y_{ij})}{\omega^2\{\mathsf{E}(Y_{ij})\}^2} = \frac{\exp(\omega^2) - 1}{\omega^2} = \frac{1 + \omega^2 + o(\omega^2) - 1}{\omega^2} = 1 + o(1),$$

so that part (iii) holds. ■

The next results require the delta method given as follows.

Lemma A.1 (The delta method van der Vaart 1998, Theorem 3.1): *Let T_ω be a sequence of random vectors or variables indexed by ω. Suppose $\omega^{-1}(T_\omega - \theta)$ is asymptotically normal $\mathsf{N}(0, \Omega)$ for $\omega \to 0$ and that g is a vector or scale valued function that is differentiable in a neighbourhood of θ with derivative $\dot{g}$. Then $\omega^{-1}\{g(T_\omega) - g(\theta)\}$ is asymptotically normal with mean zero and variance $\{\dot{g}(\theta)\}\Omega\{\dot{g}(\theta)\}'$.*

Proof of Theorem 3.3: Throughout the proof we ignore the indices i,j.

1. We show that

$$\omega^{-1}\{Y - \exp(\mu)\} = \omega^{-1}\{Y - \mathsf{E}(Y)\} + O(\omega) \tag{A1}$$

First, we add and subtract $\mathsf{E}(Y)$ term in $Y - \exp(\mu)$ to get

$$\omega^{-1}\{Y - \exp(\mu)\} = \omega^{-1}\{Y - \mathsf{E}(Y)\} + \omega^{-1}\{\mathsf{E}(Y) - \exp(\mu)\}. \tag{A2}$$

By Assumption 3.2(i), then $\mathsf{E}(Y) = \exp(\mu + \omega^2/2)$ so that the second term becomes

$$\mathcal{E}_2 = \omega^{-1}\{\mathsf{E}(Y) - \exp(\mu)\} = \omega^{-1}\exp(\mu)\{\exp(\omega^2/2) - 1\}.$$

Taylor expand the exponential function as $\exp(\omega^2/2) - 1 = \omega^2/2 + O(\omega^4)$ to get

$$\mathcal{E}_2 = \exp(\mu)\{\omega/2 + O(\omega^3)\} = O(\omega),$$

since the canonical parameter ξ is fixed, and hence μ_{ij} is fixed. The expression (A1) then follows.

2. We show that

$$\omega^{-1}\{Y\exp(-\mu) - 1\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, 1). \tag{A3}$$

Apply (A1) and divide by $\exp(\mu)$, multiply and divide by $\sqrt{\mathsf{Var}(Y)}/\omega$ and $\mathsf{E}(Y)$ to get

$$\frac{Y - \exp(\mu)}{\omega\exp(\mu)} = \frac{Y - \mathsf{E}(Y)}{\omega\exp(\mu)} + O(\omega) = \{\frac{Y - \mathsf{E}(Y)}{\sqrt{\mathsf{Var}(Y)}}\}\{\frac{\sqrt{\mathsf{Var}(Y)}}{\omega\mathsf{E}(Y)}\}\{\frac{\mathsf{E}(Y)}{\exp(\mu)}\} + O(\omega).$$

Assumption 3.2(i,iii) implies that the second and third terms converge to unity. Theorem 3.1, using Assumption 3.1. shows the first term is asymptotically normal. Dividing by $\exp(\mu)$ in numerator and denominator establishes (A3).

3. Apply the delta method in Lemma A.1 to (A3) with $T_\omega = Y\exp(-\mu)$ and $\theta = 1$ and choose $g(t) = \log(t) + \mu$, so $\dot{g}(t) = 1/t$. Then $g(T_\omega) = \log Y$ and $g(\theta) = \mu$ while $\dot{g}(\theta) = 1$ so that $\omega^{-1}(\log Y - \mu)$ is asymptotically standard normal as desired. ■

Proof of Theorem 3.4: Theorem 3.1 shows that $\{Y_{ij} - \mathsf{E}(Y_{ij})\}/\sqrt{\mathsf{Var}(Y_{ij})}$ is asymptotically standard normal. Now, Assumption 3.2(iii) shows $\mathsf{Var}(Y_{ij})/\{\omega^2\mathsf{E}^2(Y_{ij})\} \to 1$, while Assumption 3.2(i,ii) implies $\log \mathsf{E}(Y_{ij}) \to \mu_{ij}$. Combine these three results to get the desired statement. ■

Proof of Theorem 3.5: The model equation is $y_{ij} = \log Y_{ij} = X_{ij}'\xi + \varepsilon_{ij}$, see (8). Theorem 3.3, using Assumptions 3.1 and 3.2, shows that the vector of innovations $\omega^{-1}\varepsilon = \omega^{-1}(y - X\xi)$ is asymptotically standard normal as $\omega \to 0$. We can then use standard least squares distribution theory in the limit.

Recall $\hat{\xi} = (X'X)^{-1}X'y$, see (9). Substitute $y = X\xi + u$ to get

$$\omega^{-1}(\hat{\xi} - \xi) = \omega^{-1}\{(X'X)^{-1}X'(X\xi + \varepsilon) - \xi\} = (X'X)^{-1}X'(\omega^{-1}\varepsilon).$$

Since $\omega^{-1}\varepsilon \overset{D}{\to} \mathsf{N}(0, I_n)$, we have $\omega^{-1}(\hat{\xi} - \xi) \overset{D}{\to} \mathsf{N}\{0, (X'X)^{-1}\}$ as required.

The residuals in (9) can be written as $\hat{\varepsilon} = P_\perp y$, where $P_\perp = \{I_n - X(X'X)^{-1}X'\}$ is an orthogonal projection matrix so that $P_\perp = P_\perp'$ and $P_\perp^2 = P_\perp$. Inserting the model equation, this becomes $\hat{\varepsilon} = P_\perp\varepsilon$, while $P_\perp X = 0$. Since $\omega^{-1}\varepsilon \overset{D}{\to} \mathsf{N}(0, I_p)$, then $\omega^{-1}P_\perp\varepsilon \overset{D}{\to} \mathsf{N}(0, P_\perp)$, so that $\omega^{-2}s^2$ is asymptotically $\chi^2_{n-p}/(n-p)$ noting $\mathrm{tr}(P_\perp) = n - p$. Finally, $\hat{\xi}$ and s^2 are asymptotically independent, since $\hat{\xi} - \xi$ and s^2 are functions of $X'\varepsilon$ and $P_\perp\varepsilon$, while $\omega^{-1}\varepsilon$ is asymptotically standard normal, while $P_\perp X = 0$. ■

Proof of Theorem 3.8: Recall the forecast taxonomy (18), summed over $\mathcal{A}$.

The first contribution is the process error and satisfies

$$\omega^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\} = \omega^{-1} \sum_{(i,j)\in\mathcal{A}} \{Y_{ij} - \mathsf{E}(Y_{ij})\}.$$

This is a sum of independent terms, each of which is asymptotically $\mathsf{N}\{0, \exp(2\mu_{ij})\}$ by Theorem 3.4. Therefore, $\omega^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\}$ is asymptotically $\mathsf{N}(0, \varsigma^2_{\mathcal{A},process})$, where $\varsigma^2_{\mathcal{A},process} = \sum_{(i,j)\in\mathcal{A}} \exp(2\mu_{ij})$ as stated in (19), (20).

The second contribution is the estimation error from $\hat{\xi}$. Theorem 3.5 shows that as $\omega \to 0$ then $\omega^{-1}(\hat{\xi} - \xi) \overset{D}{\to} \mathsf{N}\{0, (X'X)^{-1}\}$. Apply the delta method in Lemma A.1 with $T_\omega = \hat{\xi}$ and $g(T) = \sum_{(i,j)\in\mathcal{J}} \exp(X'_{ij}\xi)$, so that $\dot{g}(T) = \sum_{(i,j)\in\mathcal{J}} \exp(X'_{ij}\xi)X'_{ij}$. Therefore, $\omega^{-1}\{\exp(X'_{ij}\hat{\xi}) - \exp(X'_{ij}\xi)\}$ is asymptotically $\mathsf{N}(0, \varsigma^2_{\mathcal{A},estimation})$, where $\varsigma^2_{\mathcal{A},estimation}$ is given in (22). Further, by continuity $\exp(\omega^2/2) \to 1$ as $\omega^2 \to 0$. In combination we arrive at (21).

The third term is the contribution from estimation error of s^2. By continuity, we get $\exp(\omega^2/2) \to 1$ as $\omega^2 \to 0$, while $\sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\xi)$ is fixed. Rewrite $s^2 = (s^2/\omega^2)\omega^2$. Since s^2/ω^2 converges in distribution by Theorem 3.5 as $\omega^2 \to 0$ then s^2 vanishes in probability. Applying the exponential function, which is a continuous mapping, yields that $\exp(s^2/2) \to 1$ in probability and so does the entire third term.

The process error and the estimation error are independent as they are based on the independent upper and lower triangles $\mathcal{J}$ and $\mathcal{I}$. Therefore, the first and second contributions to the forecast taxonomy (18) are independent, while the third contribution vanishes, so that

$$\omega^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\} \overset{\mathsf{D}}{\to} \mathsf{N}(\varsigma^2_{\mathcal{A},process} + \varsigma^2_{\mathcal{A},estimation}),$$

which is asymptotically independent of s^2. Further, s^2/ω^2 is asymptotically $\chi^2_{n-p}/(n-p)$ so that $s^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\}$ is asymptotically t_{n-p} as desired. ■

Appendix 2. Further table

Table A1. Estimates for the US casualty data for extended chain-ladder, H_{apc}.

μ_{11}	7.689	$\mu_{21}-\mu_{11}$	0.0929	$\mu_{12}-\mu_{11}$	2.076
$\Delta^2\alpha_3$	−0.133	$\Delta^2\beta_3$	−1.347	$\Delta^2\gamma_3$	0.343
$\Delta^2\alpha_4$	−0.422	$\Delta^2\beta_4$	−0.690	$\Delta^2\gamma_4$	0.044
$\Delta^2\alpha_5$	0.427	$\Delta^2\beta_5$	−0.134	$\Delta^2\gamma_5$	−0.312
$\Delta^2\alpha_6$	−0.532	$\Delta^2\beta_6$	−0.272	$\Delta^2\gamma_6$	0.170
$\Delta^2\alpha_7$	0.181	$\Delta^2\beta_7$	0.036	$\Delta^2\gamma_7$	−0.253
$\Delta^2\alpha_8$	0.177	$\Delta^2\beta_8$	−0.297	$\Delta^2\gamma_8$	0.249
$\Delta^2\alpha_9$	0.008	$\Delta^2\beta_9$	0.131	$\Delta^2\gamma_9$	0.065
$\Delta^2\alpha_{10}$	−0.118	$\Delta^2\beta_{10}$	−0.090	$\Delta^2\gamma_{10}$	−0.042
$\Delta^2\alpha_{11}$	0.119	$\Delta^2\beta_{11}$	0.219	$\Delta^2\gamma_{11}$	−0.268
$\Delta^2\alpha_{12}$	−0.471	$\Delta^2\beta_{12}$	−0.073	$\Delta^2\gamma_{12}$	0.335
$\Delta^2\alpha_{13}$	0.050	$\Delta^2\beta_{13}$	−0.003	$\Delta^2\gamma_{13}$	−0.341
$\Delta^2\alpha_{14}$	0.707	$\Delta^2\beta_{14}$	−0.321	$\Delta^2\gamma_{14}$	0.247
$\Delta^2\alpha_{15}$	0.018	$\Delta^2\beta_{15}$	0.255	$\Delta^2\gamma_{15}$	−0.010
$\Delta^2\alpha_{16}$	−0.579	$\Delta^2\beta_{16}$	0.709	$\Delta^2\gamma_{16}$	0.095
$\Delta^2\alpha_{17}$	0.436	$\Delta^2\beta_{17}$	−1.276	$\Delta^2\gamma_{17}$	−0.227
$\Delta^2\alpha_{18}$	0.031	$\Delta^2\beta_{18}$	0.984	$\Delta^2\gamma_{18}$	0.202
$\Delta^2\alpha_{19}$	−0.258	$\Delta^2\beta_{19}$	−0.463	$\Delta^2\gamma_{19}$	0.229
$\Delta^2\alpha_{20}$	0.890	$\Delta^2\beta_{20}$	0.034	$\Delta^2\gamma_{20}$	0.236
s^2	0.181	*RSS*	27.626		