Generalized Log-Normal Chain-Ladder

Di Kuang

Submitted for SA0 of
The Institute and Faculty of Actuaries

Dec 2018

Acknowledgements

I would like to express my deepest gratitude and appreciation to my supervisors Prof. Bent Nielsen, Oxford University and Dr. Pietro Parodi, Scor Insurance London, for their guidance and support throughout the project.

Abstract

The dissertation shows how the age-period-cohort model can be used to estimate, select models and make distriburion forecasts in resering. The dissertation includes a joint paper with Prof Bent Nielsen titled Generalized Log-Normal Chain-Ladder. In this paper, 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.

KEYWORDS chain-ladder, infinitely divisibility, over-dispersed Poisson, bootstrap, log-normal.

Contents

Chapter 1

Foreword

This chapter is written to ensure that this dissertation fulfils the specific passing criteria of subject SA0. This chapter references and assumes knowledge of the joint paper presented as Chapter 2 and 3 of this dissertation.

1.1 Declaration

This dissertation contains a joint paper with Prof. Bent Nielsen. My supervisor Bent Nielsen is the main contributor of the asymptotic theories of the joint paper and have guided me through out my SA0 project. I have contributed to the following areas of work independently:

- I recognized the importance of the Log-Normal model in practice and recommended to extend the asymptotic distribution forecast theories of the over-dispersed Poisson work by Harnau & Bent (2017) to the log-normal model.

- I decided what to show in the key exhibits in the empirical illustration section that are most important in practice. For example the forecast error trends by accident years, the 1 in 200 values and the standard deviation by mean values.

- I extended the current apc package for R by Nielsen (2015) to the log-normal model. I wrote all the code for the empirical and simulation studies independently.

- I collected and analysed the Lloyd's data independently. I fitted and compared the log-normal results with the over-dispersed Poisson and bootstrap results. I presented the findings to the Lloyd's actuarial team. I sought feedback from the actuaries at Lloyd's and added model selections analysis to model stability testing work for the paper.

- I searched for datasets which can be used in the paper for publication purposes. The datasets used in the paper are representations of when a log-normal model is more suitable than the traditional over-dispersed Poisson model.

- I studied Harnau and Nielsen (2017)'s paper independently and came up with an initial draft of the theories and proofs for the Generalized Log-Normal Chain-Ladder paper.

- I have coded and produced all tables and figures for the Generalized Log-Normal Chain-Ladder paper.

- I drafted an initial, 80-page version of the dissertation before sitting down with Prof. Bent Nielsen to select results and to write the main body of the paper together. I have written the first chapter of the dissertation and the conclusion of the joint paper independently.

1.2 Evaluation of achievement of research objectives

The dissertation illustrates applications of the age-period-cohort models to insurance data and produces reserve forecast distributions as set out in the approved research outline submitted to the IFoA in October 2016 (see section 3.3 in the appendix). It provides a thorough literature review of the age-period-cohort model and its application in insurance as set out in the introduction section of the joint paper.

Results from the majority of the methods of investigation in the research outline are included in this dissertation. These include generalized linear model, likelihood analysis, statistical tests, chain ladder method, bootstrap method and simulation as set out in the research outline 4.1 and best estimates and forecast distributions from the commonly used over-dispersed Poisson model, the boostrap method and the proposed log-normal model on real insurance data as set out in research outline 4.2 and 4.3.

The over-dispersed Poisson model is used frequently by practitioners but is not fully understood. The dissertation revisits the over-dispersed Poisson model and it proposed a new forecast method, the log-normal chain-ladder via an asymptotic approach. The proposed methods will improve the actuarial process which is an important objective in the research outline. I also performed analysis on the Lloyds data and presented the results to Lloyds internally however due to data confidentiality am unable to present the full results in this dissertation. A summary of findings of the datasets I have studied in this project is contained in the conclusion of the dissertation and a publicly available dataset is used in the joint paper for illustration purposes.

Investigations were also carried out using methods 4.4 and 4.5 and although they resulted in some results in this dissertation they did not end up being fully explored and hence are included as suggestions for future research. This is reasonable as they are relatively less important to the project objectives than 4.1, 4.2 and 4.3 which describe how practitioners can set their reserve in the first place. A summary of the work carried out regarding 4.4 and 4.5 is as follows:

For 4.4, I studied the working paper by Margraf & Nielsen (2018). This paper describes the new Bornhuetter-Ferguson method which allows actuaries to adjust the relative ultimates for the Poisson model. Similar linear constraints can be used in the generalized log-normal model proposed in my joint paper with Nielsen in Chapter 2. Instead of analyzing a restricted Poisson likelihood described in Margraf & Nielsen (2018), the Bornhuetter-Ferguson method can be implemented by analyzing the restricted least squares in a generalized log-normal model, which leads to t and F statistics as described in Chapter 2 of my dissertation, with an empirical example in section 2.4.1.

Section 2.4.4 of this dissertation is a result of investigation using method 4.5, in which the mean and variance of different parts of the reserving triangle are compared at the same time using the GLM package in R for both the over-dispersed Poisson model and the generalized log-normal model. This allows actuaries to select an appropriate distribution to forecast reserves. Additionally an example of interpretation of the trend based on the canonical parameterization for the apc model is given in section 2.4.1 of this dissertation. Some investigation was carried out into other parts of 4.5 including estimation of correlations between triangles, however useful results were not obtained and so do not appear in this dissertation.

1.3 Discussion of data used in this project

The reserving process starts with business plan estimates and will typically be revised quarterly until settlement. Reserve forecasts are required because there are delays between business being written, accident occurrence, accident reporting, claims handling and the final settlement. Reserve forecasts become more accurate with time as policies become more mature and more accidents related to the origin year are observed.

To model this delay, the reserving data is typically collected in a triangular format as shown in Figure 1.3.1 below. This is called the run-off triangle. Rows represent the origin years (accident years, underwriting years or reporting years depending on type of policies) and columns represent the development periods (monthly, quarterly or yearly etc.). Past information is stored in the upper triangle and the future reserves are in the lower triangle.

Development years

Origin years

£m	1	2	3	4	5	6	7	8	9	10
1	451,288	339,519	333,371	144,988	93,243	45,511	25,217	20,406	31,482	1,729
2	448,627	512,882	168,467	130,674	56,044	33,397	56,071	26,522	14,346	
3	693,574	497,737	202,272	120,753	125,046	37,154	27,608	17,864		
4	652,043	546,406	244,474	200,896	106,802	106,753	63,688			
5	566,082	503,970	217,838	145,181	165,519	91,313				
6	606,606	562,543	227,374	153,551	132,743					
7	536,976	472,525	154,205	150,564						
8	554,833	590,880	300,964							
9	537,238	701,111								
10	684,944									

Figure 1.3.1: Incremental RSA UK Motor paid triangle

Organising the data in the run-off triangle is useful as it shows how claims develop through time in different origin years. The chain-ladder method assumes claim developments are identical for different origin years. It estimates the past development rate and applies it to the future to obtain reserves.If the data is homogeneous the run-off triangle method is a useful form to present it. We can calculate the average development trend from the run off triangle and use it to project into the future. Aggregating data into the triangular format has the benefit of reducing data error due to allocation between currencies, risk groups etc.

Similar to the chain-ladder method, the age-period-cohort models incremental triangular data. It takes account of trends along the origin years, development years, as well as calendar year. The calendar year trend can be significant if inflation is prominent in the data.

A triangle free approach in Parodi (2012) has received much interest. The author argues data is lost when aggregated into the triangular format. This approach is credible for heteogenous data, for example, for large and unusual claims. In which case additional policy and claims information and expert adjustments can be made at a case by case level to ensure more accurate forecasts. For attritional claims, the traditional triangular approach could be more stable due to the aggregations. The triangle method is robust to data errors which is hard to avoid in reserving. The level of granularity is an important decision in reserving and a number of factors should be taken into consideration:

- Are data available and of sufficient quality for granular reserving? Are they correct? How can we best use the data to tell us what we want to know?

- How homogeneous are the data? Are the data we want to combine together likely to have the same trends and volatility?

- What is the purpose of the work? Do we need to report reserves at aggregate level (for example to calculate the balance sheet for an insurance company) or more granular level (for example to price a particular product or to assess performance of a particular class of business for underwriters)?

- How material is it? Would more granular reserving improve results? Larger or uncertain claims are likely to receive more attention, which may be worth the extra work.

Often actuaries uses more than one types of data in setting reserve, including paid and incurred, count and amount data.

The benefit of using incurred data is that it includes the case estimates and provides additional information than the pure paid data. The incurred data is useful when evaluating short tail classes such as property claims, in particular catastrophic losses. These classes can be verified quickly and the case estimates are generally more accurate and easy to determine. Incurred data may be less useful for long tail business such as casualty lines and financial lines, where case estimates are not so reliable and the size of the case reserves rely on the judgement of claims handlers.

The paid data records the actual payments so it involves less judgements and estimates on the ultimate amount. The incremental paid data is less likely to be negative than the incurred data as we expect curmulative payment to increase, while incurred data could have reserve releases when case estimates are over-estimated.

Claim number triangles are particularly useful for long tail business where inflation is important as claims take a long period to settle. Inflation effect can be different from the amount and number of claims. This can be modelled by a deterministic method, the average cost per claim (ACPC) method. A stochastic method, the Double Chain Ladder method by Martínez-Miranda et al. (2011) is a simulation based approach to forecast the distribution of aggregate claims.

The age-period-cohort models described in this dissertation uses trianglular data. We look at two distributions of the age-period-cohort models, the log-normal and the over-dispersed Poisson. For these distributions, paid data is more suitable than incurred data as these distributions assume incremental payments are positive. In particular, the log-normal only takes positive values, while the over-dispersed Poisson can model claims with few negative values as long as the cumulative values are positive. The frequency claims triangles are less suitable to be modelled by the log-normal and the over-dispersed Poisson distributions as they are continous distributions, with exception when the dispersion of the over-dispersed Poisson is 1. Both the log-normal and over-dispersed Poisson models are infinitely divisible, which means they are sum of many independently identically distributed distributions. In particular the underlying data generation process for the over-dispersed Poisson model is a Compound Poisson Process. The Compound Poisson Process is a summation of non-negative independently identical distributions where the number in the summation follows a Poisson distribution. Therefore they are suitable in modelling the aggregate claims data.

Although the models in this paper can be applied to both triangular data gross and net of reinsurance, it is better for gross data as generally less negative incremental claims are in the gross data. The net triangles could have some recoveries income in late development years which results in negative incremental claims. The log-normal model cannot handle negative claims and the over-dispersed Poisson model can handle some negative incremental claims as long as the row sum and column sums of the claims are positive.

In practice we cannot avoid negative incremental claims, they are generally present in many triangles, particularly net of reinsurance claims triangles and in non casualty lines of business due to salvages and subrogations. Some practitioners remove negative claims or use a shifted distribution in modelling. There have been little theorectical bases for this approach. How to deal with negative incremental claims remains an open research area for the log-normal and over-dispersed Poisson modelling.

I analyzed the high level classes of business data from Lloyd's. I have chosen the net of reinsurance data from Lloyd's so that I can compare the fitted trends by the age-period-cohort models to the Lloyd's selected trends easily. Lloyd's internal parameterisations are done on the net data as Lloyd's do not model reinsurance contracts of the whole market in detail.

The Lloyd's data is relatively stable as it is a pool of the all syndicates data. This makes it more suitable for analysis as the parameters estimates from the age-period-cohort models are reflective of underlying trends and not caused by changes in reporting or recording the data. Moreover the asymptotic theory developed for the over-dispersed Poisson model by Harnau & Nielsen (2017) and the log-normal model in this dissertation work best when the aggregate claims are large and the underlying standard deviation is small. For a given class, business placed through Lloyd's tends to be more specialist in nature, leading to more volatile claims experience. This volatility of the Lloyd's claims data contributes to less accurate forecasts from the age-period-cohort models. By aggregating the market data, it also has the benefit of limiting negative values in the triangle which are more suitable for the over-dispersed Poisson and log-normal model fits.

Due to data confidentiality, I cannot present detailed analysis results of the Lloyd's data. Hence the data analysed in the paper in Chapter 2 comes from a publically available source. There is little publicly available commercial historic loss data and much of the available data seems to come from a group of multinational Bermuda-domiciled insurers perhaps because of regulatory disclosure requirements. Many of these companies have similar profiles and XL Catlin's data is selected for use in this paper on account of their relatively large size and good quantity of historic data. The specific data used is gross paid traingulations from XL Catlin's US Casualty class. I chose the US Casualty class as it is suitable to illustrate how estimation, specification tests and distribution forecasts can be done in practice. The volatility trend forecasts for the XL Catlin are similar to those from the Lloyd's data. There is already an example of when the over-dispersed Poisson model and the bootstrap method work better than the log-normal model in Harnau & Nielsen (2017). The XL Catlin data chosen in this dissertation is an example of when the proposed log-normal model performs better than the traditional over-dispersed Poisson and bootstrap methods.

1.4 Discussion of various models

We compare log-normal forecast with the existing methods, including the Mack's method by Mack (1999), the bootstrap method by England & Verrall (1999) and England (2002) and the over-dispersed Poisson model proposed by Harnau & Nielsen (2017).

1.4.1 Mack's method

The benefit of Mack's method is that it is a formula-based method, it is easy to calculate and practictioners find it relatively stable in comparison to the bootstrap method. The disadvantage of this method is that it does not provide a full distribution of reserves. While the 1 in 200 values are very important the Mack's method only provides the first two moments of the reserve forecast errors.

1.4.2 Bootstrap method

The bootstrap method is widely used. The benefit is that it is a simulation based method which is easily understood by practitioners. The algorithm is chain ladder method based. It can be applied to a wide range of triangular data, paid and incurred, number and amount, gross and net. It offers flexibilities for practictioners to fine tune their results by using different definitions of residuals, for example a different standard deviation can be used for different development columns when calculating residuals. The disadvantage is that it is easily distored by large values of residuals when there are outliers. It is hard to justify removal of residuals to get sensible results and it produces results which only suitable for the over-dispersed Poisson model.

1.4.3 Asymptotic approach for the over-dispersed Poisson model

The asymptotic approach for the over-dispersed Poisson model recommended by Harnau and Nielsen (2017) provides a formula for the chain ladder reserve distribution which are similar to the boostrap method. The benefits of this method are that it tends to be more robust to large residuals as it does not involve calculating and randomising the residuals and that it is formula based so does not require a simulation engine. It can be easily implemented in the apc package by Nielsen (2015). The disadvantage of this is that practitioners may find it less intuitive to understand.

1.4.4 Asymptotic approach for the generalized log-normal model

The log-normal model developed in the joint paper in this dissertation provides an alternative toolset to the over-dispersed Poisson model. The benefit of the log-normal model is that it is recognized in practice as being a good choice for modelling reserves and attritional losses. The log-normal model is currently used by many practictioners, especially in simulating reserve risk in capital modeling. While the over-dispersed Poisson model is useful in modelling aggregate claims as a Compound Poisson process, the infinite divisibility property of the log-normal model proved by Thorin (1979) shows the log-normal model is just as useful in modelling aggregate claims. The asymptotic theory developed in the joint paper in this dissertation provides a formula for the log-normal forecast distribution which can be easily calculated by practitioners. We find good stability of the log-normal distribution forecasts in different reserving years in the empirical data analysis. While the over-dispersed Poisson model generally produces a decreasing volatility trend by origin years, the log-normal model produces a flatter trend. The log-normal volatilities are also generally larger in respect to their expected forecasts in comparison to the over-dispersed Poisson model.

We performed statistical tests and find that although the log-normal model is suitable to be used on some insurance data, the majority of the data are better suited to be modelled by the over-dispersed Poisson model. Out of 15 empirical data analyses I studied I found three data sets where the log-normal fits better than the over-dispersed Poisson model, two of the three datasets are casualty business. This could be due to casualty business being long tail. The volatility trend relative to the expected reserve appear to be flatter than short tail classes due to slow recognition of size and number of claims. The log-normal model assumes the standard error is proportional to the expected reserves, while the over-dispersed Poisson assumes variance proportional to the expected reserve. Therefore the log-normal model produces a flatter trend and could be more suitable for long tail classes. However specification tests show the log-normal model does not fit all casualty classes I have studied. Practitioners are therefore advised to perform the model specification tests described in the joint paper in this dissertation to test the log-normal and over-dispersed Poisson model fitness and decide what should be appropriate to use for different data.

1.5 Actuarial process

The age-period-cohort model puts the chain-ladder and bootstrap methods into a statistical framework, with a consistent process in estimation, point forecast and distribution forecast, therefore a consistent process in reserving and capital modelling can be developed. Standard statistical theories can also be applied. This allows users to perform model specification tests which provides statistical evidence of model fitness. An age-period-cohort modelling process which actuaries can follow is described below:

- Collect and check data and decide the granularity to model. If there are abnormal entries investigate source of possible errors. If there are negative developments present, consider whether to remove or adjust them for log-normal analysis or whether to use alternative data, so that the log-normal model and the over-dispersed Poisson model can fit.

- Perform model specification tests to check the log-normal and over-dispersed Poisson model fitness. This involves checking the mean and variance structure are the same in different parts of the triangle. Examples of specification tests are F-test and Bartlett test as shown in Table 2.4.5. The over-dispersed Poisson and the log-normal should cover most of the reserving classes. Select a suitable model which fits the data based on the model specification tests.

- Estimate the parameters and the associated errors of the parameter estimates using the apc package in R by Nielsen (2015) as in Table 2.4.2. Interpret trends in the parameters estimates.

- Check significance of the parameters and decide if the model benefits from dropping insignificant parameters via t-tests as in Table 2.4.2. F-statistics can also be calculated to check if the model can be reduced. An example of the F-statistics to check the signifiance of calendar year effect is shown in Table 2.4.1. Practitioners are advised to look at the results from both t-test and F-test in deciding if the model can be reduced. For example, parameters along the origin years can be jointly significant, but not individually significant as shown in the empirical example in the joint paper. In this case practitioners can consider to drop some α parameters. These test results can be computed via the apc package by Nielsen (2015) in R.

- Produce forecast distribution statistics and inspect them graphically as shown in Table 2.4.3 and Figure 2.4.1. The over-dispersed Poisson forecast distribution is coded in the apc package by Nielsen (2015). This package is planned to be extended to include log-normal forecast distribution forecasts in the near future. The forecast distribution is a t-distribution and it can be fed into capital modelling to simulate reserve risks.

Parameterization of reserve risk in practice requires actuaries' to employ judgement when making data adjustments, selections, and assessing the appropriateness of model results. In particular factors such as changes in terms and conditions, mix of business and other factors not captured in the data may not be well reflected in the estimated trend straight from the model which forms part of this analysis.

1.6 Contribution to actuarial science

This dissertation contributes to actuarial science by putting the lognormal model which is already in use by practitioners on a more sound theoretical footing. It derives theoretical results which can be used to select a suitable model in reserving exercises, which is a regulatory requirement that has been found to be difficult to implement in practice using established techniques. The age-period-cohort model framework presented here also provides a good theoretical foundation for future research to extend the basic chain ladder method, and could be useful in pricing and capital modelling as well as reserving.

1.7 Contribution to actuarial practice

This dissertation contributes to actuarial practice as it provides an alternative set of tools to model reserves. It exhibits empirical examples based on real datasets that illustrate how the log-normal forecasts based on the asymptotic theory developed in this dissertation work and how the results compare to the existing models. It suggests a process which practitioners can follow to produce estimates, forecasts and model selections (as mentioned in 1.6) and also includes specimen R code to allow quick real-life initialisation of the methods.

Chapter 2

Generalized Log-Normal Chain-Ladder paper

This chapter presents a joint paper with Prof. Bent Nielsen which was submitted to the Scandinavian Actuarial Journal in April 2018.

2.1 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 and Verrall (1999) and England (2002) is commonly used, but it does not always produce satisfactory results. Recently, Harnau and Nielsen (2017) 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 of Harnau and Nielsen (2017) 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 consists of a reserving triangle of aggregate amounts that have been paid with some delay in respect to portfolios of insurances. Table 2.1.1 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.

The chain-ladder is maximum likelihood in a Poisson model. This is useful for estimation and point forecasting. Martínez Miranda, Nielsen and Nielsen (2015) have developed a theory for inference and distribution forecasting in such a Poisson model in order to analyze 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, Pentikäinen and Pesonen (1984, §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 analyze and even harder to convolute. England and 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 and Nielsen (2017) 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, keep-

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1997	2185	13908	44704	56445	67313	62830	72619	42511	32246	51257	11774	21726	10926	4763	3580	4777	1070	1807	824	1288
1998	3004	17478	49564	55090	75119	66759	76212	62311	31510	15483	23970	8321	15027	3247	8756	14364	3967	3858	4643	
1999	5690	28971	55352	63830	71528	73549	72159	37275	38797	27264	28651	14102	8061	17292	10850	10732	4611	4608		
2000	9035	29666	47086	41100	58533	80538	70521	40192	27613	13791	17738	20259	12123	6473	3922	3825	3082			
2001	7924	38961	41069	64760	64069	61135	62109	52702	36100	18648	32572	17751	18347	10895	2974	5828				
2002	7285	25867	44375	58199	61245	48661	57238	29667	34557	8560	12604	8683	9660	4687	1889					
2003	3017	22966	62909	54143	72216	58050	29522	25245	19974	16039	8083	9594	3291	2016						
2004	1752	25338	56419	75381	64677	58121	38339	21342	14446	13459	6364	6326	6185							
2005	1181	24571	66321	65515	62151	43727	29785	23981	12365	12704	12451	8272								
2006	1706	13203	40759	57844	48205	50461	27801	21222	14449	10876	8979									
2007	623	14485	27715	52243	60190	45100	31092	22731	19950	18016										
2008	338	6254	24473	32314	35698	25849	30407	15335	15697											
2009	255	3842	14086	26177	27713	15087	17085	12520												
2010	258	7426	22459	28665	32847	28479	24096													
2011	1139	10300	19750	32722	41701	29904														
2012	381	5671	34139	33735	33191															
2013	605	11242	24025	32777																
2014	1091	9970	31410																	
2015	1221	8374																		
2016	2458																			

Table 2.1.1: XL Group, US casualty, gross paid and reported loss and allocated loss adjustment expense in 1000 USD.

ing 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, Nielsen and Nielsen (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 analyzed 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 analyzed by for instance Kremer (1982), Renshaw (1989), Verrall (1991, 1994), Doray (1996) and England and 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, Nielsen and Nielsen (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 variable 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 and Nielsen (2017) 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 and Zehnwirth (2000), and Kuang, Nielsen and Nielsen (2008a,b,2011). The variance could be different in subgroups of the triangle as pointed out by Hertig (1985). Barlett (1937) proposed a test for this problem. Recently, Harnau (2017) has adapted that test to the traditional chain-ladder. We extend this to the generalized log-normal model. Figure 2.1.1 shows key references of this joint paper.

We illustrate the new methods using a casualty reserving triangle from XL Group (2017) as shown in Table 2.1.1. 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 2.1.1. We find that that the asymptotic results give good approximations in finite samples. The asymptotic 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 §2.2 we review the well known log-normal models for reserving. In §2.3 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 §2.4, while §2.5 provides the simulation study. Finally, we discuss directions for future research in §2.6. All proofs of theorems are provided in an Appendix.

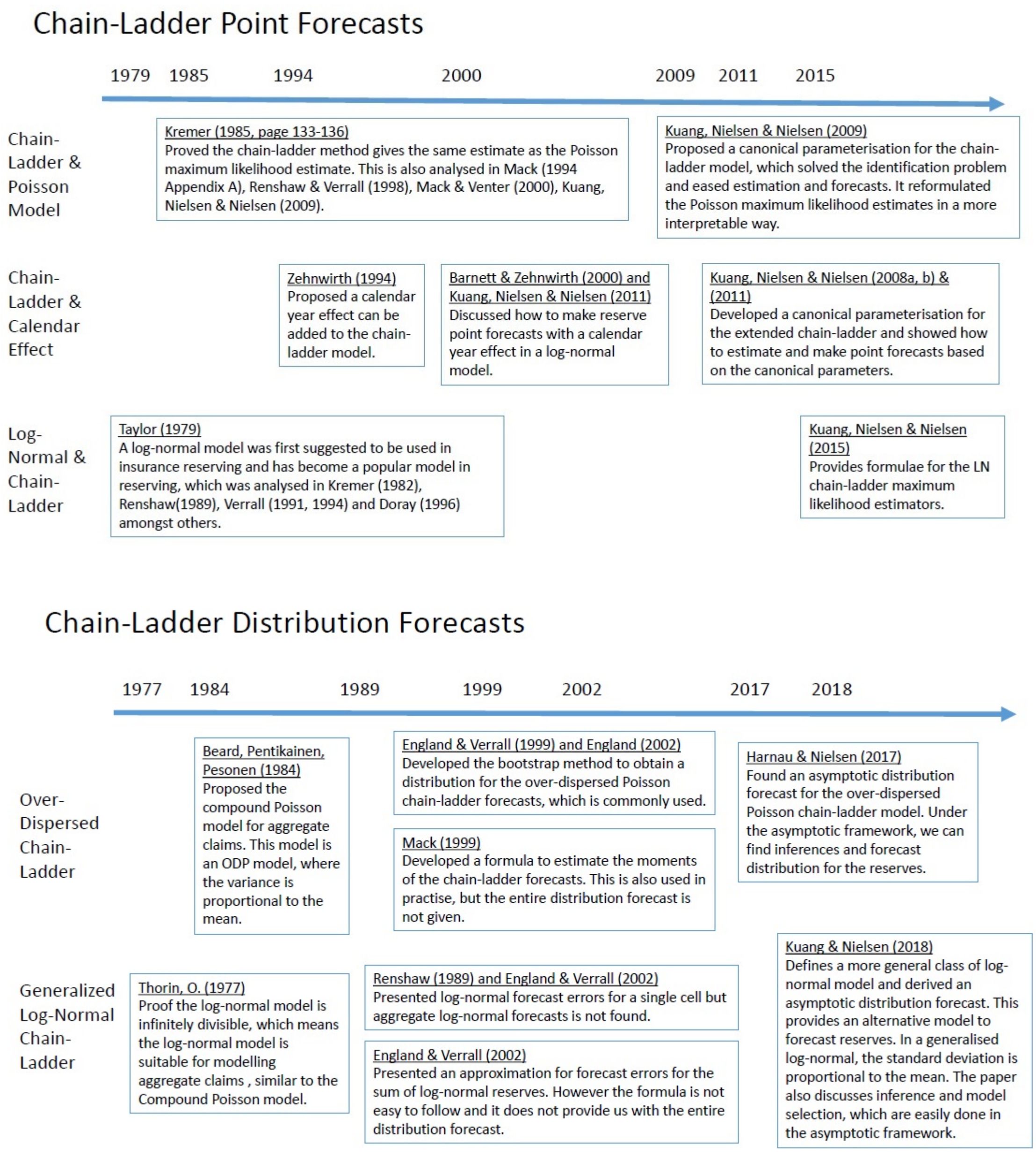


Figure 2.1.1: Key references on the over-dispersed Poisson and log-normal chain-ladder forecasts

2.2 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.

2.2.1 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

$$\mathcal{I} = \{i, j : \ i \text{ and } j \text{ belong to } (1, \ldots, k) \text{ with } i + j - 1 = 1, \ldots, k\}. \tag{2.2.1}$$

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, Nielsen and Nielsen (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

$$\mathcal{J} = \{i, j : \ i \text{ and } j \text{ belong to } (1, \ldots, k) \text{ with } i + j - 1 = k+1, \ldots, 2k-1\}. \tag{2.2.2}$$

2.2.2 Model

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

$$\mu_{ij} = \alpha_i + \beta_j + \delta. \tag{2.2.3}$$

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 2.2.1 ***log-normal model***. *The array Y_{ij}, $i, j \in \mathcal{I}$, satisfies that the variables $y_{ij} = \log Y_{ij}$ are independent normal $\mathsf{N}(\mu_{ij}, \omega^2)$ distributed, where the predictor is given by (2.2.3)*

The parametrisation presented in (2.2.3) 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 parametrisation of Kuang, Nielsen and Nielsen (2009, 2015) so that the model becomes a regular exponential family with freely varying parameters. The canonical parameter is

$$\xi = \{\mu_{11}, \Delta\alpha_2, \ldots, \Delta\alpha_k, \Delta\beta_2, \ldots, \Delta\beta_k\}', \tag{2.2.4}$$

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

$$\mu_{ij} = \mu_{11} + \sum_{\ell=2}^{i} \Delta\alpha_\ell + \sum_{\ell=2}^{j} \Delta\beta_\ell = X_{ij}'\xi, \tag{2.2.5}$$

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

$$X_{ij}' = \{1, 1_{(2\leq i)}, \ldots, 1_{(k\leq i)}, 1_{(2\leq j)}, \ldots, 1_{(k\leq j)}\}, \tag{2.2.6}$$

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

2.2.3 Statistical analysis

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

$$\ell_{\log \mathsf{N}}(\xi, \omega^2) = -\frac{n}{2}\log(2\pi\omega^2) - \frac{1}{2\omega^2}\sum_{i,j\in\mathcal{I}}(y_{ij} - X_{ij}'\xi)^2. \tag{2.2.7}$$

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

$$y = X\xi + \varepsilon. \tag{2.2.8}$$

The least squares estimator for ξ and the residuals are then

$$\hat{\xi} = (X'X)^{-1}X'y, \qquad \hat{\varepsilon}_{ij} = y_{ij} - X_{ij}'\hat{\xi}. \tag{2.2.9}$$

while the variance ω^2 is estimated by

$$s^2 = \frac{RSS}{n-p} \quad \text{where} \quad RSS = \sum_{i,j\in\mathcal{I}} \hat{\varepsilon}_{ij}^2. \tag{2.2.10}$$

Kuang, Nielsen and Nielsen (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 and Nielsen (2007), so that

$$\hat{\xi} \overset{\mathsf{D}}{=} \mathsf{N}\{\xi, \omega^2(X'X)^{-1}\}, \qquad s^2 \overset{\mathsf{D}}{=} \chi^2_{n-p}/(n-p). \tag{2.2.11}$$

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

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

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

$$F = \frac{\{RSS_H - RSS\}/(p - p_H)}{RSS/(n-p)} \overset{\mathsf{D}}{=} \mathsf{F}(p - p_H, n - p_H). \tag{2.2.12}$$

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 Margraf and Nielsen (2018). This is analyzed by restricted least squares which also leads to t and F statistics.

2.2.4 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 $\mathcal{J}$, as well as the predictor for the upper triangle $\mathcal{I}$, can be formed as

$$\tilde{Y}_{ij} = \exp(X_{ij}'\hat{\xi} + \hat{\omega}^2/2), \tag{2.2.13}$$

We will also be interested in distribution forecasting. However, 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 and Nielsen (2017).

2.2.5 Extending with a calendar effect

It is common to extend the chain-ladder parametrization with a calendar effect, so that linear predictor in (2.2.3) becomes

$$\mu_{ij,apc} = \alpha_i + \beta_j + \gamma_{i+j-1} + \delta, \tag{2.2.14}$$

where $i+j-1$ is the calendar year corresponding to accident year i and development year j. This model has been suggesting in insurance by Zehnwirth (1994). Similar models have been used in a variety of displines 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 problems. The canonical parameter solution of Kuang, Nielsen and Nielsen (2008a) is to write $\mu_{ij,apc} = X'_{ij,apc}\xi_{apc}$ where, with $h(i,s) = \max(i - s + 1, 0)$, we have

$$\begin{aligned} \xi_{apc} &= (\mu_{11}, \nu_a, \nu_c, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_k, \Delta^2\beta_3, \ldots, \Delta^2\beta_k, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_k)', && (2.2.15) \\ X_{ij,apc} &= \{1, i-1, j-1, h(i,3), \ldots, h(i,k), h(j,3), \ldots, h(j,k), \\ &\qquad h(i+j-1,3), \ldots, h(i+j-1,k)\}. && (2.2.16) \end{aligned}$$

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

This model can be analyzed 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 (2.2.9) 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 (2.2.12), 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 identication problem, see Kuang, Nielsen and Nielsen (2008b, 2011).

2.3 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 and Nielsen (2017) to provide distribution forecasts for the generalized log-normal model.

2.3.1 Assumptions and first properties

The infinitely divisible setup of Harnau and Nielsen (2017, §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 $X_1, \ldots, X_m$ 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 need 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 2.3.1 ***Infinite divisibility***. *The array Y_{ij}, $i, j \in \mathcal{I}$, satisfies*
(i) Y_{ij} are independent distributed, non-negative and infinitely divisible;
(ii) asymptotically, 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 analyzing characteristic function and exploiting the Lévy-Kintchine formula for infinitely divisible distributions.

Theorem 2.3.1 ***(Harnau and Nielsen, 2017, Theorem 1)*** *Suppose Assumption 2.3.1 is satisfied. Then*

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

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 (2.2.5), (2.2.6) or an extended chain-ladder model with a calendar effect.

Assumption 2.3.2 ***The generalized log-normal chain-ladder model***. *The array Y_{ij}, $i, j \in \mathcal{I}$, satisfies Assumptions 2.3.1 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 (2.2.7);*
(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 2.2.1 is indeed of the generalized log-normal model.

Theorem 2.3.2 *Consider the log-normal model of Assumption 2.2.1. Suppose the dimension of the array $\mathcal{I}$ is fixed as $\omega^2 \to 0$. Then Assumptions 2.3.1, 2.3.2 are satisfied.*

A first consequence of the generalized log-normal model is that Theorem 2.3.1 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 2.3.3 is useful in deriving the inference in Theorem 2.3.5 and estimation error for forecasts in Theorem 2.3.8 in later sections.

Theorem 2.3.3 *Suppose Assumptions 2.3.1, 2.3.2 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 $i, j \in \mathcal{I}$ then the standardized y_{ij} are asymptotically independent standard normal.

We will need to reformulate the Central Limit Theorem 2.3.1 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 2.3.4 is useful in deriving the process error for forecasts in Theorem 2.3.8 later.

Theorem 2.3.4 *Suppose Assumptions 2.3.1, 2.3.2 are satisfied. Then, as $\omega^2 \to 0$,*

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

Note that Y_{ij} over $i, j \in \mathcal{I}$ are assumed independent.

2.3.2 Inference

We check that the inferential results for the log-normal model, described in §2.2.3, 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 2.3.5 *Consider the generalized log-normal model defined by Assumptions 2.3.1, 2.3.2 and the least squares estimators (2.2.9). Then, as $\omega^2 \to 0$,*

$$\begin{aligned} \omega^{-1}(\hat{\xi} - \xi) &\overset{\mathsf{D}}{\to} \mathsf{N}\{0, (X'X)^{-1}\}, \\ \omega^{-2}s^2 &\overset{\mathsf{D}}{\to} \chi^2_{n-p}/(n-p). \end{aligned}$$

The estimators $\hat{\xi}$ and s^2 convergent jointly and are asymptotically independent.

We can derive inference for of the estimator $\hat{\xi}$ using asymptotic t distribution. The proof follows Theorem 2.3.5 and the Continuous Mapping Theorem.

Theorem 2.3.6 *Consider the generalized log-normal model, defined by Assumptions 2.3.1, 2.3.2. 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 (2.2.12) from the classical normal model. The proof is similar to Theorem 4 of Harnau and Nielsen (2017).

Theorem 2.3.7 *Consider the generalized log-normal model, defined by Assumptions 2.3.1, 2.3.2 with three types of linear predictor:*
the extended chain-ladder model parametrised by $\xi_{apc} \in \mathbb{R}^{p_{apc}}$ *in (2.2.15);*
the chain-ladder model parametrised by $\xi \in \mathbb{R}^p$ *in (2.2.4); and*
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* $\omega \to 0$,

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

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

2.3.3 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 a cash flow of 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 (2.2.13), while the overall point forecast is

$$\tilde{Y}_{\mathcal{A}} = \sum_{(i,j)\in\mathcal{A}} \tilde{Y}_{ij} = \sum_{(i,j)\in\mathcal{A}} \exp(X'_{ij}\hat{\xi} + s^2/2) \qquad (2.3.1)$$

To find the forecast error we expand

$$\begin{aligned} Y_{ij} - \tilde{Y}_{ij} = \{Y_{ij} - \mathsf{E}(Y_{ij})\} - \exp(\omega^2/2)\{\exp(X'_{ij}\hat{\xi}) - \exp(X'_{ij}\xi)\} \\ + \{\exp(\omega^2/2) - \exp(s^2/2)\}\exp(X'_{ij}\hat{\xi}), \quad (2.3.2) \end{aligned}$$

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 (2.3.2) is the process error. When extending Theorem 2.3.4 to the lower triangle $\mathcal{J}$ we will get

$$\omega^{-1}\{Y_{\mathcal{A}} - \mathsf{E}(Y_{\mathcal{A}})\} \xrightarrow{\mathsf{D}} \mathsf{N}(0, \varsigma^2_{\mathcal{A},process}), \qquad (2.3.3)$$

where

$$\varsigma^2_{\mathcal{A},process} = \sum_{i,j\in\mathcal{A}} \exp(2X'_{ij}\xi) \tag{2.3.4}$$

The second term in (2.3.2) is the estimation error for the canonical parameter ξ. From Theorem 2.3.5 we will be able to derive

$$\omega^{-1}\exp(\omega^2/2)\{\exp(X'_{ij}\hat{\xi}) - \exp(X'_{ij}\xi)\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, \varsigma^2_{\mathcal{A},estimation}), \tag{2.3.5}$$

where

$$\varsigma^2_{\mathcal{A},estimation} = \{\sum_{i,j\in\mathcal{A}} \exp(X'_{ij}\xi)X'_{ij}\}(X'X)^{-1}\{\sum_{i,j\in\mathcal{A}} \exp(X'_{ij}\xi)X_{ij}\}. \tag{2.3.6}$$

The third term in (2.3.2) vanishes asymptotically. We will estimate ω^2 by s^2, which turns the asymptotic normal distributions into t-distribution. 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 2.3.8 *Suppose the generalized log-normal model defined by Assumptions 2.3.1, 2.3.2 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

$$r^2_{\mathcal{A},process} = \sum_{i,j\in\mathcal{A}} \exp(2X'_{ij}\hat{\xi}), \tag{2.3.7}$$

$$r^2_{\mathcal{A},estimation} = \{\sum_{i,j\in\mathcal{A}} \exp(X'_{ij}\hat{\xi})X'_{ij}\}(X'X)^{-1}\{\sum_{i,j\in\mathcal{A}} \exp(X'_{ij}\hat{\xi})X_{ij}\}. \tag{2.3.8}$$

Thus, the distribution forecast is

$$\tilde{Y}_{\mathcal{A}} + \{\hat{\omega}^2(r^2_{\mathcal{A},process} + r^2_{\mathcal{A},estimation})\}^{1/2}\mathsf{t}_{n-q}. \tag{2.3.9}$$

2.3.4 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 (2017) 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.

Suppose the triangle $\mathcal{I}$ can be divided into two or more groups as indicated in Figure 2.3.1. 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 seperately 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 seperately 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

$$RSS_\ell = \sum_{i,j \in \mathcal{I}_\ell} (y_{ij} - \hat{\mu}_{ij,\ell})^2, \qquad s_\ell^2 = \frac{1}{df_\ell} RSS_\ell. \tag{2.3.10}$$

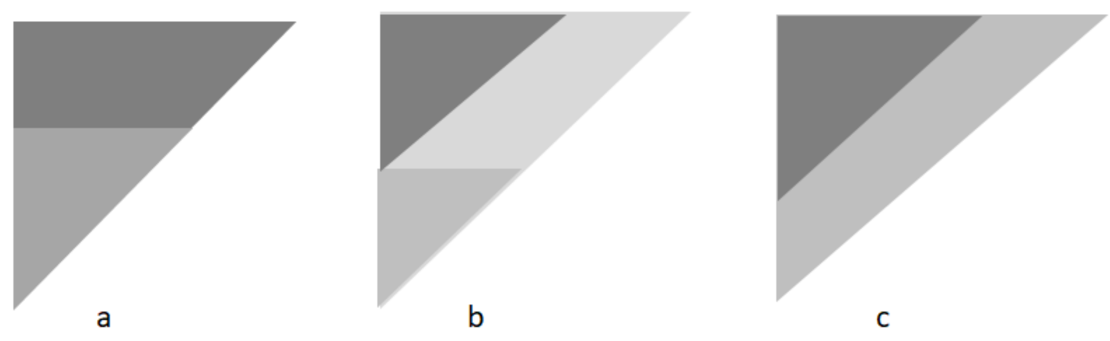


Figure 2.3.1: Examples of dividing triangles in two parts

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

$$F^\omega = s_2^2/s_1^2 \overset{\mathsf{D}}{=} \mathsf{F}_{n_2-p_2, n_1-p_1}. \tag{2.3.11}$$

In the generalized log-normal the F-distribution can be shown to be valid asymptotically. Harnau (2017) 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 $s_1^2, \ldots, s_m^2$ are independent scaled χ^2 variables. Bartlett found the likelihood for this χ^2 model. Under the hypothesis the common variance is estimated by

$$\bar{s}^2 = \frac{1}{df.} \sum_{\ell=1}^m RSS_\ell, \qquad \text{where} \qquad df. = \sum_{\ell=1}^m df_\ell = n - \sum_{\ell=1}^m p_\ell, \tag{2.3.12}$$

while the likelihood ratio test statistic for the hypothesis is

$$LR^{\omega} = df_{\cdot} \log(\bar{s}^2) - \sum_{\ell=1}^{m} df_{\ell} \log(s_{\ell}^2). \tag{2.3.13}$$

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

$$\frac{LR^{\omega}}{C} \approx \chi^2_{m-1} \qquad \text{where} \qquad C = 1 + \frac{1}{3(m-1)}\left(\sum_{\ell=1}^{m} \frac{1}{df_{\ell}} - \frac{1}{df_{\cdot}}\right). \tag{2.3.14}$$

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 exceptionelly well in very small samples; see the simulations by Harnau (2017). Once again the Bartlett test (2.3.13) will be applicable in the generalized log-normal model, which can be proved by following the proof of Harnau (2017).

In practice, we can fit seperate log-normal models to each group, that is $y_{ij\ell}$ is assumed $\mathsf{N}(\mu_{ij\ell}, \omega_{\ell}^2)$. 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.

2.4 Empirical illustration

We apply the theory to the insurance run-off triangle shown in Table 2.1.1. All R (2017) code is given in the supplementary material. We use the R packages `apc`, 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.

2.4.1 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 2.4.1 shows an analysis of variance. This conforms with the exact distribution theory in §2.2.3 and the asymptotic distribution theory in Theorems 2.3.5, 2.3.7 in §2.3.2.

sub	$-2\log \mathsf{L}$	df_{sub}	$\mathrm{F}_{sub,apc}$	p	$\mathrm{F}_{sub,ac}$	p
apc	170.00	153				
ac	179.87	171	0.41	0.984		
ad	258.57	189	2.23	0.000	4.32	0.000

Table 2.4.1: Analysis of variance for the US casualty data

Other model reductions can also be tested similarly, such as the age-period (ap) model which was proposed by Zehnwirth (1994). For illustration purpose we consider the three commonly considered models for reserves. First, we test the chain-ladder model (ac for age-cohort) against the extended chain-ladder model (apc for age-period-cohort) with $\mathsf{p} = 0.984$. 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 $\mathsf{p} = 0.000$ and $\mathsf{p} = 0.000$ when testing against the apc and ac models respectively. 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 2.4.2 shows the estimated parameters for the log-normal model with chain-ladder structure (ac). We report standard errors se_t following Theorem 2.3.6. They are the same for $\Delta\alpha$ and $\Delta\beta$ due to symetry 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. The parameters with * are when the absolute t-values are greater than 1.97, that is outside the 2-sided 5% test. In these cases we conclude the parameter are insignificant. 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 2.1.1. At the same time many the accident year effects $\Delta\alpha$ are not individually significant, although they are jointly significant as seen in Table 2.4.1. It is suggested that practitioners reduce their models by dropping of the most insignificant parameter iteratively until all parameters are significant. All parameters are retained in the analysis in this paper so that the forecasts can be easily compared with those from the usual chain ladder and bootstrap methods. The signs of the $\Delta\alpha$'s match the relative increase or decrease of the amounts seen in the rows of Table 2.1.1.

In Appendix 3.2 we present a further Table 3.2.1 with estimates. These are the estimated parameters for the log-normal model with an extended chain-ladder structure (apc) as in §2.2.5. 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.

	estimate	t-value			estimate	t-value	
μ_{11}	7.66	55.59	*				
$\Delta\alpha_2$	0.29	17.02	*	$\Delta\beta_2$	2.27	2.16	*
$\Delta\alpha_3$	0.16	6.84	*	$\Delta\beta_3$	0.93	1.20	
$\Delta\alpha_4$	-0.27	1.68		$\Delta\beta_4$	0.24	-1.90	
$\Delta\alpha_5$	0.15	0.62		$\Delta\beta_5$	0.09	1.04	
$\Delta\alpha_6$	-0.37	-1.19		$\Delta\beta_6$	-0.18	-2.52	*
$\Delta\alpha_7$	-0.20	-0.94		$\Delta\beta_7$	-0.14	-1.30	
$\Delta\alpha_8$	-0.01	-2.69	*	$\Delta\beta_8$	-0.43	-0.06	
$\Delta\alpha_9$	-0.01	-1.82		$\Delta\beta_9$	-0.30	-0.03	
$\Delta\alpha_{10}$	-0.13	-2.32	*	$\Delta\beta_{10}$	-0.40	-0.77	
$\Delta\alpha_{11}$	-0.02	-1.05		$\Delta\beta_{11}$	-0.19	-0.12	
$\Delta\alpha_{12}$	-0.47	-1.28		$\Delta\beta_{12}$	-0.24	-2.49	*
$\Delta\alpha_{13}$	-0.44	-1.30		$\Delta\beta_{13}$	-0.26	-2.18	*
$\Delta\alpha_{14}$	0.30	-2.60	*	$\Delta\beta_{14}$	-0.56	1.38	
$\Delta\alpha_{15}$	0.31	-1.32		$\Delta\beta_{15}$	-0.30	1.35	
$\Delta\alpha_{16}$	-0.27	1.62		$\Delta\beta_{16}$	0.41	-1.08	
$\Delta\alpha_{17}$	0.14	-3.23	*	$\Delta\beta_{17}$	-0.90	0.51	
$\Delta\alpha_{18}$	0.20	0.37		$\Delta\beta_{18}$	0.12	0.64	
$\Delta\alpha_{19}$	-0.09	-1.01		$\Delta\beta_{19}$	-0.38	-0.25	
$\Delta\alpha_{20}$	0.87	-0.54		$\Delta\beta_{20}$	-0.27	1.72	
s^2	0.17			RSS	28.96		

Table 2.4.2: Estimates for the US casualty data for the log-normal chain-ladder (ac).

2.4.2 Distribution forecasting

Table 2.4.3 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

	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	11699	0.26	1.66	10975	0.54	2.41	10846	0.56	2.90
6	13717	0.24	1.64	14941	0.44	2.14	14767	0.45	2.50
7	14344	0.22	1.58	18337	0.39	2.01	18147	0.40	2.29
8	18377	0.21	1.54	24487	0.34	1.87	24233	0.35	2.09
9	25488	0.21	1.54	31876	0.29	1.76	31607	0.30	1.93
10	30525	0.20	1.53	35567	0.28	1.72	35270	0.28	1.87
11	40078	0.20	1.53	48595	0.24	1.63	48176	0.25	1.73
12	32680	0.20	1.53	42027	0.26	1.68	41659	0.27	1.80
13	28509	0.21	1.54	37114	0.28	1.74	36814	0.29	1.88
14	51761	0.21	1.55	66977	0.22	1.58	66554	0.23	1.69
15	98748	0.22	1.58	102982	0.20	1.51	102282	0.20	1.59
16	100331	0.23	1.60	136647	0.19	1.51	135880	0.20	1.59
17	149813	0.24	1.64	164318	0.22	1.56	163500	0.22	1.68
18	221550	0.26	1.69	218874	0.25	1.66	218115	0.26	1.83
19	229481	0.30	1.79	166120	0.49	2.29	166431	0.51	2.84
20	575343	0.41	2.06	337001	0.94	3.46	353628	1.03	4.91
total	1656586	0.16	1.42	1469605	0.23	1.60	1480500	0.26	1.95

Table 2.4.3: Forecasting for the US casualty data using the generalized log-normal, the over-dispersed Poisson model and the bootstrap. The bootstrap simulation is based on 10^5 repetitions.

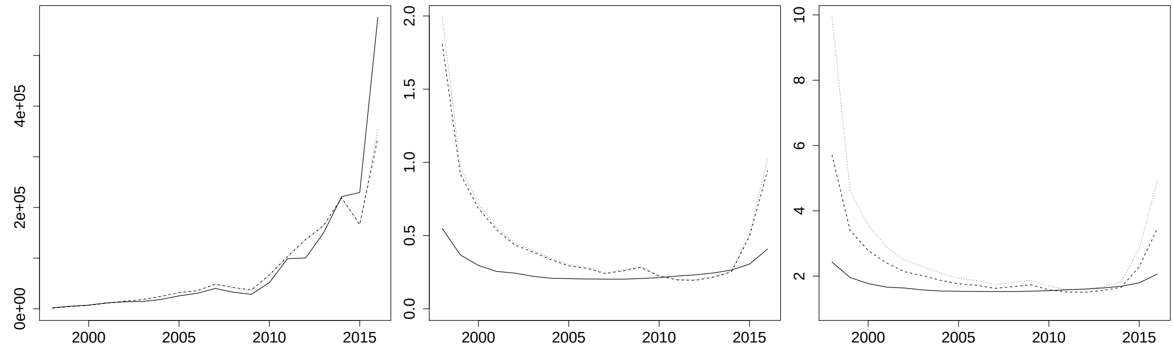
forecast values (99.5%/Res).

For the generalized log-normal chain-ladder model we use the asymptotic distribution forecast in (2.3.9). For the over-dispersed Poisson model we use the asymptotic distribution forecasts from Harnau and Nielsen (2017, equation 11). For the bootstrap we use the ChainLadder package by Gesmann et al (2005), based on the method described in England (2002). We apply 10^5 bootstrap draws using the gamma option.

Table 2.4.3 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 additional factor $\exp(s^2/2) = \exp(0.17/2) = 1.09$ 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 2.4.1 shows the trends of the reserve and standard error and 99.5% quantile



(a) Reserve (b) se/Reserve (c) 99.5%/Reserve

Figure 2.4.1: Illustration of the forecasts in Table 2.4.3 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.

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. 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.

2.4.3 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 2.4.4 shows standard error and 99.5% over reserve ratios. It has 9 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 5 accident years based on the full triangle. These numbers are the same as those in Table 2.4.3. 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 5 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 2.1.1, 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.

Full triangle			Leave 1 out			Leave 2 out		
generalized log-normal								
i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$	i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$	i	$\frac{se}{Res}$	$\frac{99.5\%}{Res}$
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

Table 2.4.4: Recursive forecasting for the US casualty data in the latest 5 accident years. The bootstrap simulation is based on 10^5 repetitions.

2.4.4 Model selection

We now apply the specification test outlined in §2.3.4 for the log-normal model and in Harnau (2017) for the over-dispersed Poisson model. For the tests we split the data triangle of Table 2.1.1 as outlined in Figure 2.3.1:

(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 sub-group. We then compute the Bartlett test statistic LR^{ω}/C from (2.3.14) for a common variance across groups. Given a common variance we also compute an F-statistic for common chain-ladder structure in the mean. The Bartlett test statistics LR^{ω}/C follows a chi-square distribution and the F-statistic is F-distributed. When the relevant p-values for these statistics are small, for example, less than 2.5%, it indicates the mean and variance structure are differ in the triangle.

For each of the generalized log-normal and over-dispersed Poisson model we are conducting 6 tests. When choosing 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 and Nielsen (2007, §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.

	generalized log-normal				over-dispersed Poisson			
	Bartlett test		F test		Bartlett test		F test	
splits	LR^{ω}/C	p	F	p	LR^{ω}/C	p	F	p
(a)	6.29	1.2%	1.34	3.0%	11.68	0.1%	1.43	1.2%
(b)	4.70	9.5%	1.55	0.5%	11.63	0.3%	2.50	0.0%
(c)	1.12	29.1%	1.33	3.7%	15.07	0.0%	1.24	9.3%

Table 2.4.5: Bartlett tests for common dispersion and F tests for common mean parameters.

Starting with the log-normal model we see that there is only moderate evidence against model. The worst cases are that variance differs across the (a) split and the chain-ladder structure differs across the (b) split. The log-normal fails in 2 out of 6 tests at the 2.5% level, while the over-dispersed Poisson model is rejected by 5 out of 6 tests. Therefore the log-normal model is preferred here over the over-dispersed Poisson model. The variance differ across the (a) split could be caused by underwriting cycle. Claims prior to 2003 are generally more volatile than claims after 2003 across the insurance market. The mean difference suggests the development pattern has changed in the recent years. Practitioners can fit two models to model the different mean structure.

2.5 Simulation

In Theorems 2.3.7 and 2.3.8 we presented asymptotic results for inference and distribution forecasting. We now apply simulation to investigate the quality of these asymptotic approximations.

2.5.1 Test statistic

We assess the finite sample performance of the F-tests proposed in Theorem 2.3.7 and applied in Table 2.4.1. 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 2.3.7 also applies for infinitely divisible distributions that are not log-normal but satisfy Assumptions 2.3.1 and 2.3.2. 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 2.3.7 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 (2.2.5) where the parameters μ_{ij} are chosen to match those of Table 2.4.2. We also choose σ^2 to match the estimate s^2 from Table 2.4.2, 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 3.2.1. In all cases we draw 10^5 repetitions.

	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%

Table 2.5.1: Simulated performance of F test based on 10^5 draws. The Monte Carlo standard error less than 0.2%.

The size is the type I error, which is the probability of incorrectly rejecting the null hypothesis, that is, calendar year trend is not present. The smaller the size, the better the performance of the F-test. We note that the simulated size is correct apart from Monte Carlo standard error, where Monte Carlo standard error calculation formula can be found in equation (18.1.9), page 272 of Hendry & Nielsen (2017). This is because the $\mathsf{F}(18, 153)$-distribution is exactly the same as the confidence level of the test under the null hypothesis. We are operating on the log-scale and simulate normal variables so that standard regression theory applies.

Under the alternative we simulate power, unity minus type II error. This is the probability of correctly accepting the calendar year trend being useful. The higher the power the better the F-tests. The simulations show that the power increases for shrinking variance $v^2\omega^2$ for F-tests at different confidence levels. For example, at the 5% confidence level, the F-test as shown in Table 2.4.1 works correctly just under 3 out of 10 times and it improves to working correctly over 9 out of 10 times when the standard deviation of the data reduce by 50%. When selecting an appropriate confidence level practitioners should consider the implied size and power of the resultant test, ensuring both are sufficient to allow meaningful test results.

When simulating the power of the F-test in Table 2.5.1, the test is not exact, but asymptotically F-distributed with shrinking variance. Therefore the power increases with smaller v. 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 2.3.7, and the non-centrality diverges in a similar way even though n is fixed.

2.5.2 Forecasting

We assess the finite sample performance of the asymptotic distribution forecasts proposed in Theorem 2.3.8 and applied in Table 2.4.3. These asymptotic distribution forecasts are compared to the over-dispersed Poisson forecast of Harnau and Nielsen (2017) and the bootstrap of England and 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 2.4.2. 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 and Ashe (1983) data, see also Harnau and Nielsen (2017, 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 2.3.8 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 2.3.8 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 2.5.2 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 2.5.2 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 and Nielsen (2017) and the bootstrap of England and 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 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 (2.3.9) 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 2.5.2 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 2.5.3 we repeat the simulation exercise for the Taylor and Ashe (1983) data. For these data we repeated the empirical exercise of §2.4, 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

v		Moments Mean	SE	Quantiles 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
v	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
v	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

Table 2.5.2: Simulation performance of distribution forecasts for the US casualty data. Results in USD. The study is based on 10^5 repetitions, and for each simulated upper triangle, the bootstrap is based on 999 simulations.

		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
v	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
v	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

Table 2.5.3: Simulation performance of distribution forecasts for the data used in Taylor & Ashe (1983) Results. The study is based on 10^5 repetitions, and for each simulated upper triangle, the bootstrap is based on 999 simulations.

Harnau and Nielsen (2017, Table 2). As before we simulate a log-normal distribution with parameters equal to the estimates from the data.

Table 2.5.3 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 $v = 1/2$.

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 and Verrall (1999) and England (2002) along with the recent asymptotic theory of Harnau and Nielsen (2017).

Actuaries 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 ladder 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 (2017) for the traditional chain ladder is a beginning of that research agenda.

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.

Chapter 3

Appendix

This Chapter shows proofs of Theorems and additional tables for the joint paper in Chapter 2, the approved reserve outline and the refereces for this dissertation.

3.1 Proofs of Theorems

Proof of Theorem 2.3.2. Recall the following results. A log-normally distributed variable Y_{ij} is positive, hence non-negative. It is infinitely divisible as shown by Thorin (1977). The first three cumulants are

$$\mathsf{E}(Y_{ij}) = \exp(\mu_{ij} + \omega^2/2), \tag{3.1.1}$$
$$\mathsf{Var}(Y_{ij}) = \exp(2\mu_{ij} + \omega^2)\{\exp(\omega^2) - 1\}, \tag{3.1.2}$$
$$\frac{\mathsf{E}\{Y_{ij} - \mathsf{E}(Y_{ij})\}^3}{\{\mathsf{Var}(Y_{ij})\}^{3/2}} = \{\exp(\omega^2) - 2\}^{1/2}\{\exp(\omega^2) + 2\}, \tag{3.1.3}$$

see Johnson, Kotz and Balakrishnan (1994, equations 14.8a, 14.8b and 14.9a).

The log-normal distribution is a non-degenerate and non-negative divisible distribution, see Thorin (1977) and

$$\begin{aligned} skew(Y) &= \frac{E(Y - E(Y))^3}{\sqrt{Var(Y)}^3} = \frac{\exp(3\omega^2) - 3\exp(\omega^2) + 2}{(exp(\omega^2) - 1)^{3/2}} \\ &= \frac{1 + 3\omega^2 + \frac{1}{2}9\omega^4 - 3\left(1 + \omega^2 + \frac{\omega^4}{2}\right) + 2 + O(\omega^6)}{(1 + \omega^2 - 1)^{3/2}} \\ &= \frac{\left(\frac{9}{2} - \frac{3}{2}\right)\omega^4 + O(\omega^6)}{\omega^3} = 3\omega + O(\omega^3) \to 0. \end{aligned}$$

as $\omega \to 0$. □

The next results require the delta method given as follows.

Lemma 3.1.1 ***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 2.3.3. Throughout the proof we ignore the indices i, j.

1. We show that

$$\omega^{-1}\{Y - \exp(\mu)\} = \omega^{-1}\{Y - \mathsf{E}(Y)\} + \mathrm{O}(\omega) \tag{3.1.4}$$

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

$$\omega^{-1}\{Y - \exp(\mu)\} = \omega^{-1}\{Y - \mathsf{E}(Y)\} + \omega^{-1}\{\mathsf{E}(Y) - \exp(\mu)\}. \tag{3.1.5}$$

By Assumption 2.3.2(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 $\exp(\omega^2/2) - 1 = \omega^2/2 + \mathrm{O}(\omega^4)$ to get

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

since the canonical parameter ξ is fixed, and hence μ_{ij} is fixed. The expression (3.1.4) then follows.

2. We show that

$$\omega^{-1}\{Y\exp(-\mu)-1\} \xrightarrow{\mathsf{D}} \mathsf{N}(0,1). \tag{3.1.6}$$

Apply (3.1.4) and divide by $\exp(\mu)$, multiply and divide by $\sqrt{\mathsf{Var}(Y)}/\omega$ and $\mathsf{E}(Y)$ to get

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

Assumption 2.3.2(i,iii) implies that the second and third terms converge to unity. Theorem 2.3.1, using Assumption 2.3.1. shows the first term is asymptotically normal. Dividing by $\exp(\mu)$ in numerator and denominator establishes (3.1.6).

3. Apply the delta method in Lemma 3.1.1 to (3.1.6) 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 2.3.4. Theorem 2.3.1 shows that $\{Y_{ij}-\mathsf{E}(Y_{ij})\}/\sqrt{\mathsf{Var}(Y_{ij})}$ is asymptotically standard normal. Now, Assumption 2.3.2(iii) shows $\mathsf{Var}(Y_{ij})/\{\omega^2\mathsf{E}^2(Y_{ij})\}\to 1$, while Assumption 2.3.2$(i,ii)$ implies $\log\mathsf{E}(Y_{ij})\to\mu_{ij}$. Combine these three results to get the desired statement. □

Proof of Theorem 2.3.5. The model equation is $y_{ij}=\log Y_{ij}=X_{ij}'\xi+\varepsilon_{ij}$, see (2.2.8). Theorem 2.3.3, using Assumptions 2.3.1, 2.3.2, 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 (2.2.9). 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 $\omega^{-1}\varepsilon\xrightarrow{D}\mathsf{N}(0,I_n)$, we have $(\omega^{-1}(\hat\xi-\xi)\xrightarrow{D}\mathsf{N}\{0,(X'X)^{-1}\}$ as required.

The residuals in (2.2.9) 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 $\omega^{-1}\varepsilon\xrightarrow{D}\mathsf{N}(0,I_p)$, then $\omega^{-1}P_\perp\varepsilon\xrightarrow{D}\mathsf{N}(0,P_\perp)$, 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 2.3.8. Recall the forecast taxonomy (2.3.2), summed over $\mathcal{A}$.

The first contribution is the process error and satisfies

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

This is a sum of independent terms, each of which is asymptotically $\mathsf{N}\{0,\exp(2\mu_{ij})\}$ by Theorem 2.3.4. Therefore, $\omega^{-1}\{Y_{\mathcal{A}}-E(Y_{\mathcal{A}})\}$ 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 (2.3.3), (2.3.4).

The second contribution is the estimation error from $\hat{\xi}$. Theorem 2.3.5 shows that as $\omega \to 0$ then $\omega^{-1}(\hat{\xi} - \xi) \xrightarrow{D} \mathsf{N}\{0, (X'X)^{-1}\}$. Apply the delta method in Lemma 3.1.1 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 (2.3.6). Further, by continuity $\exp(\omega^2/2) \to 1$ as $\omega^2 \to 0$. In combination we arrive at (2.3.5).

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 2.3.5 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 (2.3.2) are independent, while the third contribution vanishes, so that

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

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

3.2 Further table

μ_{11}	7.69	$\mu_{21}-\mu_{11}$	0.09	$\mu_{12}-\mu_{11}$	2.08
$\Delta^2\alpha_3$	-0.13	$\Delta^2\beta_3$	-1.35	$\Delta^2\gamma_3$	0.34
$\Delta^2\alpha_4$	-0.42	$\Delta^2\beta_4$	-0.69	$\Delta^2\gamma_4$	0.04
$\Delta^2\alpha_5$	0.43	$\Delta^2\beta_5$	-0.13	$\Delta^2\gamma_5$	-0.31
$\Delta^2\alpha_6$	-0.53	$\Delta^2\beta_6$	-0.27	$\Delta^2\gamma_6$	0.17
$\Delta^2\alpha_7$	0.18	$\Delta^2\beta_7$	0.04	$\Delta^2\gamma_7$	-0.25
$\Delta^2\alpha_8$	0.18	$\Delta^2\beta_8$	-0.30	$\Delta^2\gamma_8$	0.25
$\Delta^2\alpha_9$	0.01	$\Delta^2\beta_9$	0.13	$\Delta^2\gamma_9$	0.07
$\Delta^2\alpha_{10}$	-0.12	$\Delta^2\beta_{10}$	-0.09	$\Delta^2\gamma_{10}$	-0.04
$\Delta^2\alpha_{11}$	0.12	$\Delta^2\beta_{11}$	0.22	$\Delta^2\gamma_{11}$	-0.27
$\Delta^2\alpha_{12}$	-0.47	$\Delta^2\beta_{12}$	-0.07	$\Delta^2\gamma_{12}$	0.34
$\Delta^2\alpha_{13}$	0.05	$\Delta^2\beta_{13}$	-0.00	$\Delta^2\gamma_{13}$	-0.34
$\Delta^2\alpha_{14}$	0.71	$\Delta^2\beta_{14}$	-0.32	$\Delta^2\gamma_{14}$	0.25
$\Delta^2\alpha_{15}$	0.02	$\Delta^2\beta_{15}$	0.26	$\Delta^2\gamma_{15}$	-0.01
$\Delta^2\alpha_{16}$	-0.58	$\Delta^2\beta_{16}$	0.71	$\Delta^2\gamma_{16}$	0.10
$\Delta^2\alpha_{17}$	0.44	$\Delta^2\beta_{17}$	-1.28	$\Delta^2\gamma_{17}$	-0.23
$\Delta^2\alpha_{18}$	0.03	$\Delta^2\beta_{18}$	0.98	$\Delta^2\gamma_{18}$	0.20
$\Delta^2\alpha_{19}$	-0.26	$\Delta^2\beta_{19}$	-0.46	$\Delta^2\gamma_{19}$	0.23
$\Delta^2\alpha_{20}$	0.89	$\Delta^2\beta_{20}$	0.03	$\Delta^2\gamma_{20}$	0.24
s^2	0.18	RSS	27.63		

Table 3.2.1: Estimates for the US casualty data for extended chain-ladder, H_{apc}.

3.3 Approved Research Outline

The following is the approved research outline submitted to the IFoA in October 2016.

SA0 Research Outline

Research Title: Application of Age Period Cohort Model in General Insurance Reserving

1 Indication of the research topic

I propose a research project for SA0 based on D.Phil. study at University of Oxford and five published papers jointly with my Oxford supervisor Bent Nielsen and second supervisor Jens Nielsen from Cass Business School on the topic of application of Age Period Cohort (APC) model in general insurance reserving. My research will be testing the APC model, including some recent developments on real data sets from the London market. This includes looking at parameter estimates, forecast distributions and correlations. Comparisons will be made with the traditional chain ladder, Bornhuetter-Ferguson method and the Bootstrapping method.

2 Potential benefits of the results

Potential benefits of the results will provide actuaries an alternative toolset to the simple chain ladder, which is statistically sound, and more generalised and easily implemented. The study will also demonstrate based on real data when APC model performs better than the current standard reserving techniques. The study is original in that although the APC model is widely used in epidemiology and demography study, it has not been applied to insurance data. My study will turn the APC model into a useful reserving tool and to show how it provides useful outputs required by actuaries.

3 Reason for studying it

The Reserve is the amount of money insurance company put aside for future liability. Reserving is a significant workload for an insurance company. Solvency II requires actuaries to calculate the distribution of forecasts using adequate statistical techniques, which should be consistent with the methods used to calculate best estimates. It also requires actuaries to estimate correlations to ensure diversification and dependency structure being appropriate in their internal model. The standard techniques used in the London Market are basic chain ladder for best estimate and bootstrapping for the forecast distribution. These techniques seem to produce reasonable results and are easy to implement. However, there is no theoretical basis for these techniques; best estimate, forecast distribution and correlation are not always done in a consistent manner.

The APC model can be written in the Generalized Linear Model framework which allows drawing statistical inference for the underlying data via likelihood analysis. My study will apply APC model in insurance reserving. The study will be based on the London market data from Lloyd's. It shall show a number of applications of the APC model on reserving, including incorporating with prior information in setting best estimate reserves such as the Bornhuetter Ferguson method, forecasting reserve distribution, estimation of correlations between business, comparison of data trends. This provides an alternative toolset to the current standard reserving techniques. The study will also compare the APC model outputs with the traditional techniques and shed light on types of data which the APC model performs well.

4 Methods of investigation

4.1 APC model

The main model I shall use in my study is the APC model. The APC model is similar to an extended chain ladder method in the sense that the underlying data have three trends as in an insurance claims triangle. The cohort represents year of account (or accident year), age represents claims development time and period represents calendar year effect. My papers [8] & [9] discovered a useful way to parameterise the APC model, which is a key for easily interpreting the estimates and making forecasts.

The Statistical methods I shall use in my study include the APC model, the generalized linear model, likelihood analysis, time series, statistical tests, chain ladder method, BF method, bootstrap method and simulation.

4.2 Best estimate from APC model

Estimation method is maximum likelihood analysis. For a two dimensional Age-Cohort (AC) model, I have found analytical solution to the estimators, see ref[3] & [8]. In particular, when assuming claims are (over-dispersed) Poisson distributed, the AC model produces exactly the same best estimate of reserve as a simple chain ladder.

For a three dimensional APC model, I shall use the APC and/or GLM package in statistical software R (see ref[1] & [2]) to obtain estimates of the parameters. This incorporates with a time series along the Period trend, provides forecasts for reserves. The Period trend captures changes in claims handling, changes in legislation, changes in reserving basis, changes in claims inflation and changes in policy terms and conditions, ref[5] discusses how to make robust forecasts using different time series models for the calendar year trend in order to come up with a robust reserve estimate.

4.3 Forecast Distribution from APC model

For the forecast distribution of reserves based on APC, I will use a recently developed asymptotic theory developed by Bent Nielsen. This involves an asymptotic forecast distribution assuming a sampling scheme based on infinite divisibility, which is based on an F statistics. The forecast distribution incorporates both estimation error and process error and will be compared with the bootstrapping distribution forecast. The study will also consider cases when outliers and missing data point presents in the underlying triangle data. I shall assess the forecast distribution by comparing the actual data, forecast bias and standard errors.

4.4 Bornhuetter-Ferguson (BF) method on the APC model

I will also demonstrate implementation of a variation of Bornhuetter-Ferguson (BF) method to the APC, which is also developed recently by Bent Nielsen, which in turn builds on ref[4]. The usual BF method works out the reserve by using prior ultimate information for each accident year, typically a prior loss ratio given by an underwriter. The reserve for accident year I based on BF method is:

Reserve_I = PriorUltimate_I x (1-1 / F_(k+2-I)),

where F_(k+2-I) is cumulative development factor from current year (k+2-I) to ultimate year k. The idea is that we only need to take information of relative ultimate, for example, U_I / U_1. Suppose the business volume remains the same, this would be the percentage of movement of loss ratios in accident year I compare to year 1.

4.5 Compare data trends on the APC model

The next part of my study will be comparing multiple reserving triangles. This will be done by estimating the parameters of multiple triangles at the same time via the APC (and/or) GLM package in R. This will be an entirely new method. We can then compare the reserving triangles by value of APC estimators and corresponding t statistics. Correlation between estimators will also be produced, which allows calculation of correlation between reserves between triangles.

5 Time scale of the project

I plan to work on this project over two years' part time:

- 6month to review literature and to set up codes to reproduce results in published paper s
- 1 year to analyse data and draw conclusions (2-3 months on each of the bullet points in section 3)
- 6 months to write up the dessertation

3.4 References

Barnett, G. and 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. and Pesonen, E. (1984) *Risk Theory*, 3rd edn. 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.

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

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

England, P.D. and 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. and Concina, F. (2015) Chainladder: Statistical methods and models for claims reserving in general insurance. `cran.R-project.org/package=ChainLadder`.

Harnau, J. and Nielsen, B. (2017) Over-dispersed age-period-cohort models. Nuffield College Discussion Paper. To appear in *Journal of the American Statistical Association*.

Hendry, D.F. and 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. and Balakrishnan, N. (1994) *Continuous Univariate Distributions* volume 1, 2nd edn. New York: Wiley.

Kremer, E. (1982) IBNR-Claims and the Two-way model of ANOVA. *Scandinavian Actuarial Journal*, 47–55.

Kremer, E. (1985) *Einführung in die Versicherungsmathematik*. Göttingen: Vandenhoeck & Ruprecht.

Kuang, D., Nielsen B. and 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. and 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. and Nielsen J.P. (2009) chain-ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* 4, 105-121.

Kuang, D., Nielsen B. and Nielsen J.P. (2011) Forecasting in an extended chain-ladder-type model. *Journal of Risk and Insurance* 78, 345–359.

Kuang, D., Nielsen B. and Nielsen J.P. (2015) The geometric chain-ladder. *Scandinavian Actuarial Journal*, 278–300.

Mack, T. (1991) A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, 21, 93-109.

Mack, T. (1999) The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. *ASTIN Bulletin* 29, 361–366.

Mack, T. and Venter, G. (2000) A comparison of stochastic models that reproduce the Chain Ladder estimates. Insurance: Mathematics and Economics 26, 101-107.

Margraf, C., and Nielsen, B. (2018) A likelihood approach to Bornhuetter-Ferguson analysis. Nuffield College Discussion Paper.

Martínez-Miranda, M.D., Nielsen J.P. and Verrall R. (2012) Double chain Ladder. ASTIN Bulletin, 42(1), pp. 59-76

Martínez-Miranda, M.D., Nielsen, B. and 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.

Parodi, P. (2012) Triangle-free reserving: A non-traditional protocol for estimating reserves and reserve uncertainty. Proceedings of GIRO 2012.

R Core Team (2017) 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.

Renshaw A. E. and Verrall R. J. (1998). A stochastic model underlying the chainladder technique. *British Actuarial Journal* 4, 903923.

Taylor, G. C. (1979) Statistical Testing of a Non-Life Insurance Model. Proceedings Actuarial Sciences Institute, Act. Wetemschappen, Katholieke Univ. Leuven, Belgium.

Taylor, G. C. and 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, Spring 1994*, 393–446.

XL Group (2017) *2016 Global Loss Triangles*
`http://phx.corporate-ir.net/phoenix.zhtml?c=73041&p=irol-financialreports`

Zehnwirth, B. (1994) Probabilistic development factors with applications to loss reserve variability, prediction intervals, and risk based capital. *Casualty Actuarial Society Forum, Spring 1994*, 447–605.