Technische Universität München
Department of Mathematics
Chair of Mathematical Statistics

Reserve Estimation and Analysis with Generalized Additive Models for Location, Scale and Shape

Master's Thesis

by

Florian Schewe

Supervisor:	Prof. Claudia Czado, Ph.D.
Advisor:	Jakob Stöber, M.Sc. (hons.)
Submission Date:	28.11.2012

Ich erkläre hiermit, dass ich diese Masterarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Garching, den 28. November 2012

Acknowledgments

The writing of this thesis has been one of the most significant academic challenges I have ever had to face. Without the support of several people this study would not have been completed.

I would like to express my deepest gratitude to Prof. Claudia Czado, Ph.D. for giving me the opportunity to write this master's thesis at the Chair of Mathematical Statistics. Her guidance and support has been well appreciated and I am thankful for her contributions of time and ideas.

I would like to thank Jakob Stöber, M.Sc. (hons.) for his great enthusiasm, excellent support, profounded knowledge and availability at any time during the last months. His findings, advices and ideas played a major role in the preparation of this thesis.

I would like to thank my friends, classmates and colleagues for their support during my studies and making the last years enjoyable and exciting.

I would like to thank my girlfriend for her patience, backing and care in the last couple of months.

Finally, I would like to thank my family, particularly my parents for their endless support, encouragement and love throughout my whole life.

Contents

1. Introduction to Reserving **3**
1.1. The Necessity of Reserving 3
1.2. Loss Triangles 4
1.3. Reserving 6
1.4. A Deterministic Approach: Chain Ladder Method 7
1.5. Mack's Stochastic Model for the Chain Ladder Method 10
1.6. A Generalized Linear Model for the Chain Ladder Method 17
1.7. Bootstrapping 23

2. Introduction to GAMLSS **33**
2.1. The Idea 33
2.2. The Model 34

3. Distributions **37**
3.1. Gaussian Distribution 37
3.2. t-Distribution 38
3.3. Zero-Inflated Distributions 40
3.3.1. Zero-Inflated t-Distribution 41
3.4. Skew Exponential Distribution 42
3.4.1. Skew Gaussian Distribution 44
3.5. Summary 46

4. Application of GAMLSS **47**
4.1. Modeling the Location Parameter 49
4.2. Model Fitting 51
4.3. Results 64

5. Comparison of Models **73**
5.1. Accident Year Effects 73
5.2. Calendar Year Effects 79
5.3. Dependencies 81

6. Future Claims Analysis **89**
6.1. Line of Business 1 91
6.2. Line of Business 2, 3, 4 and 5 95

7. Reserve Estimation and Dependence Analysis **99**
7.1. Line of Business 3 100
7.2. Line of Business 5 102
7.3. Other Lines of Business 105
7.4. Joint Reserve Estimation 105
7.5. Portfolio Reserve 109

8. Conclusion and Outlook **113**

A. Inflation and Dependence Models **115**
A.1. Inflation 115
A.2. Dependence 117

B. Exhibits per Line of Business **125**
B.1. Line of Business 1 125
B.1.1. Cash Flow Analysis 125
B.1.2. Latest Model & Reserving 131
B.2. Line of Business 2 133
B.2.1. Cash Flow Analysis 133
B.2.2. Latest Model & Reserving 139
B.3. Line of Business 3 141
B.3.1. Cash Flow Analysis 141
B.3.2. Latest Model & Reserving 147
B.4. Line of Business 4 149
B.4.1. Cash Flow Analysis 149
B.4.2. Latest Model & Reserving 155
B.5. Line of Business 5 157
B.5.1. Cash Flow Analysis 157
B.5.2. Latest Model & Reserving 163

List of Figures

3.1. Density plot for standard Gaussian and standardized t-distribution . . . 38
3.2. Density plots of non-standardized t-distributions with different degrees of freedom ν or scale parameter σ . . . 39
3.3. Examples of zero-inflated Gaussian distributions . . . 41
3.4. Examples of skew exponential power type 2 distributions . . . 43
3.5. Influence of the different parameters on the SEP1 distribution . . . 45
3.6. Density plot of skew normal distributions . . . 46

4.1. Development of PtP ratios for LoB 6. Although PtP ratios seem to be almost constant for higher lags in (a) they are not as one can see in (b) . 50
4.2. Development of PtP ratios among accident year for the first four development lags . . . 51
4.3. Both plots show a very bad fit for the ZIG model . . . 54
4.4. Residual plots of the ZITF 1 model show better a fit than for the ZIG 1 model . . . 55
4.5. Residual plots of the ZITF 2.1 model with fixed degree of freedom at $\nu = 2.1$ 57
4.6. Residual plots of the ZITF 2.2 model . . . 57
4.7. Residuals plots of the ZITF 2.3 model . . . 59
4.8. Zeroes appear only for later accident years . . . 60
4.9. Residuals plots of the ZITF 2.4 model . . . 61
4.10. Residuals arranged in triangular form help to identify outliers . . . 62
4.11. Both plots show a good fit for ZITF 2.5 . . . 64
4.12. Estimated values for μ for all accident years but 1993 (solid line) and for accident year 1993 (dashed line) fit well . . . 65
4.13. Plot of fitted values for σ on log-scale . . . 66
4.14. Fitted values $\hat{\tau}_k$ and observed probabilities z_k . . . 67
4.15. Empirical and estimated means and variances for model ZITF 2.5 . . . 68
4.16. ZIG 1 . . . 70
4.17. ZITF 1 . . . 70
4.18. ZITF 2.1 . . . 70
4.19. ZITF 2.2 . . . 71
4.20. ZITF 2.3 . . . 71
4.21. ZITF 2.4 . . . 72
4.22. ZITF 2.5 . . . 72

5.1. Strong non-linear trend in both data sets . . . 75
5.2. Variation increases with increasing accident years . . . 78

5.3. The Gaussian distribution has a much standard deviation for accident year 9 than for accident year 5 . . . 79
5.4. A simple scatterplot shows a strong functional relation between both lines of business . . . 82
5.5. Q-Q plot shows no lack of fit for Personal Auto GLM, but plot against the index does . . . 83
5.6. Residual plots for Personal Auto GAMLSS . . . 84
5.7. Residual plots for Commercial Auto GLM . . . 85
5.8. Residual plots for Commercial Auto GAMLSS . . . 85
5.9. Scatterplots of residuals from both approaches . . . 86

6.1. Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 1 . . . 91
6.2. Predictive distributions for LoB 1 . . . 92
6.3. Chain ladder estimates, empirical means and empirical medians from GAMLSS for LoB 3 . . . 95
6.4. Observed age-to-age factors 1-2 (a) and PtP ratios for development lag 2 (b) for line of business 3 . . . 96

7.1. Ultimate loss estimates for both methods (LoB 3) . . . 100
7.2. Estimated densities for line of business 3 . . . 101
7.3. Ultimate Loss estimates for both methods (LoB 5) . . . 103
7.4. Estimated densities for line of business 5 . . . 104
7.5. Scatterplot and Kendall's tau's of residuals . . . 108
7.6. Ultimate loss estimates for both methods of the portfolio . . . 110
7.7. Estimated densities for portfolio reserve . . . 111

A.1.1 Residuals Plots for GAMLSS applied to Paid-to-Premium ratios . . . 116
A.2.1 GLM and GAMLSS for Personal Auto . . . 122
A.2.2 GLM and GAMLSS for Commercial Auto . . . 123

B.1.1 Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 1 . . . 125
B.1.2 Residuals plots for LoB 1 (CYear 2005 & 2006) . . . 127
B.1.3 Residuals plots for LoB 1 (CYear 2007 & 2008) . . . 128
B.1.4 Residuals plots for LoB 1 (CYear 2009 & 2010) . . . 129
B.1.5 Predictive distributions for LoB 1 . . . 130
B.1.6 Residual plots for Line of Business 1 . . . 131
B.1.7 Ultimate loss estimates and estimated total reserve densities (LoB 1) . . 132
B.2.1 Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 2 . . . 133
B.2.2 Residuals plots for LoB 2 (CYear 2005 & 2006) . . . 135
B.2.3 Residuals plots for LoB 2 (CYear 2007 & 2008) . . . 136
B.2.4 Residuals plots for LoB 2 (CYear 2009 & 2010) . . . 137
B.2.5 Predictive distributions for LoB 2 . . . 138

B.2.6 Residual plots for Line of Business 2 139
B.2.7 Ultimate loss estimates and estimated total reserve densities (LoB 2) . . 140
B.3.1 Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 3 141
B.3.2 Residuals plots for LoB 3 (CYear 2005 & 2006) 143
B.3.3 Residuals plots for LoB 3 (CYear 2007 & 2008) 144
B.3.4 Residuals plots for LoB 3 (CYear 2009 & 2010) 145
B.3.5 Predictive distributions for LoB 3 146
B.3.6 Residual Plots for Line of Business 3 147
B.3.7 Ultimate loss estimates and estimated total reserve densities (LoB 3) . . 148
B.4.1 Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 4 149
B.4.2 Residuals plots for LoB 4 (CYear 2005 & 2006) 151
B.4.3 Residuals plots for LoB 4 (CYear 2007 & 2008) 152
B.4.4 Residuals plots for LoB 4 (CYear 2009 & 2010) 153
B.4.5 Predictive distributions for LoB 4 154
B.4.6 Residual plots for Line of Business 4 155
B.4.7 Ultimate loss estimates and estimated total reserve densities (LoB 4) . . 156
B.5.1 Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 5 157
B.5.2 Residuals plots for LoB 5 (CYear 2005 & 2006) 159
B.5.3 Residuals plots for LoB 5 (CYear 2007 & 2008) 160
B.5.4 Residuals plots for LoB 5 (CYear 2009 & 2010) 161
B.5.5 Predictive distributions for LoB 5 162
B.5.6 Residual plots for Line of Business 5 163
B.5.7 Ultimate loss estimates and estimated total reserve densities (LoB 5) . . 164

List of Tables

1.1. General triangle structure 4
1.2. Example of triangles with incremental and cumulative losses 5
1.3. Cumulative paid claims triangle 8
1.4. Cumulative and incremental loss triangle 21
1.5. Chain Ladder estimates (cumulative losses) 21
1.6. GLM estimates (incremental losses) 22
1.7. 3x3 loss triangle with cumulative payments 26
1.8. Step 1 27
1.9. Step 2 27
1.10. Step 3 27
1.11. Step 4 28

4.1. Triangles with incremental and cumulative earned premiums 48
4.2. Example how the index i has to be read (3x3 triangle) 55
4.3. All models for LoB 6 69

5.1. First two columns of a loss triangle 73
5.2. Premiums P_i 74
5.3. Age-to-Age factors $F_{i,1}$ and PtP ratios $PtP_{i,k}$ 74
5.4. First two columns of a loss triangle 76
5.5. Premiums P_i 77
5.6. Age-to-Age factors $F_{i,1}$ and PtP ratios $PtP_{i,k}$, $k = 1, 2$ 77
5.7. Cash Flow for calendar year 6 (bold incremental paid losses) 80
5.8. GAMLSS estimates are closer to the observed ones than chain ladder estimates 80

6.1. Number of available observations. 90

7.1. LoB 3 Total reserve statistics 102
7.2. LoB 5 total reserve statistics 104
7.3. Portfolio total reserve statistics 111

A.2.1 Commercial Auto (incremental paid losses) 118
A.2.2 Personal Auto (incremental paid losses) 118

B.1.1 All six models for LoB 1 126
B.1.2 LoB 1 total reserve statistics 131
B.2.1 All six models for LoB 2 134

B.2.2LoB 2 total reserve statistics 139
B.3.1All six models for LoB 3 142
B.3.2LoB 3 total reserve statistics 147
B.4.1All six models for LoB 4 150
B.4.2LoB 4 total reserve statistics 155
B.5.1All six models for LoB 5. 158
B.5.2LoB 5 total reserve statistics 163

Notation

Within this thesis we use the following notation and abbreviations:

$S_{i,k}$	incremental payment for accident year i at development lag k
$C_{i,k}$	cumulative payment for accident year i until development lag k
$\mathcal{S}$	loss triangle with incremental payments
$\mathcal{C}$	loss triangle with cumulative payments
R_i	reserve for accident year i
R	total reserve
$PtP_{i,k}$	Paid-to-Premium ratio for accident year i at development lag k
$F_{i,k}$	age-to-age factor for accident year i, for development from lag k to $k+1$
$\hat{\cdot}$	estimate of $\cdot$
$\hat{f}_k^{CL}$	chain ladder factor for development k to $k+1$
$\hat{C}_{i,k}^{CL}$	chain ladder estimate of cumulative payment $C_{i,k}$
X, Y	random variables
x, y	realizations of random variables X and Y, respectively
$\mathbb{E}[X]$	mean of X
$\mathbb{V}\mathrm{ar}[X]$	variance X
$\mathrm{Cov}[X, Y]$	covariance of X and Y
$\mathrm{MSE}[\cdot]$	mean squared error of $\cdot$
GLM	Generalized Linear Model
GAMLSS	Generalized Additive Model for Location, Scale and Shape
$\mathcal{N}$, NO	Gaussian Distribution
TF	t-Distribution
SEP1	Skew Exponential Power Type 1 Distribution
ODP	Over-Dispersed Poisson Distribution
ZIG	Zero-Inflated Gaussian Distribution
ZITF	Zero-Inflated t-Distribution

Preface

For non-life insurance companies, estimating reserves, known as 'reserving', is an essential and recurring task. Reserves are money put back the insurance company to pay future obligations. In general, future obligations are unknown and many methods to estimate them have been established in the last decades. The probably most famous method is the chain ladder method. Originated as a purely deterministic algorithm, it's simplicity and estimation power let it become the standard estimation technique. With the chain ladder method, estimation of ultimate losses and hence of reserves can easily and quickly be done with standard computer software.

But providing no measures of accuracy of estimation, soon a need for more sophisticated, stochastic models emerged. Kremer (1982) introduced a log-normal model for incremental payments, but Mack (1994) showed that this model is not exactly equivalent to the chain ladder method. Instead he introduced a distribution free model which yields the same reserve estimates as the chain ladder method and which allows to estimate a measure of accuracy of estimation; the mean squared error. Renshaw and Verrall (1994) casted the chain ladder method into the framework of generalized linear models with an over-dispersed Poisson model for incremental payments and based upon Kremer's findings. They could show that reserve estimates from this model coincide with estimates from the chain ladder method and Mack's distribution free model. But Mack and Venter (2000) showed that beside some other differences the mean squared error of the over-dispersed Poisson model does not agree with the mean squared error of Mack's distribution free model. Hence the models are not equivalent and a discussion about which is the better model emerged.

Beside this extensions to both models were presented to improve estimation results. Like for a portfolio of shares, there is the question if correlation of different lines of business in an insurance portfolio is present and how it affects reserve estimates. To name some of them, Braun (2004) introduced a way to model such dependencies using a multivariate chain ladder approach. More research on this was conducted by Merz and Wüthrich (2008), when they considered the prediction error of another version of the multivariate chain ladder method. Merz and Wüthrich (2009) combined the chain ladder method with an additive loss reserving method to account for inhomogeneous sub portfolios.

While these approaches model dependence among the original payments, we follow an approach introduced by Shi and Frees (2011). They fit generalized linear models to loss triangles and analyzed dependencies among the residuals of those models. However, we show that generalized linear models are not suitable for loss triangles in many situations.

Especially the assumption of variance homogeneity is violated and thus dependence modeling among residuals is not applicable. We fit generalized additive models for location, scale and shape (GAMLSS), introduced by Rigby and Stasinopoulos (2001, 2005) to loss triangles and point out their advantage over generalized linear models and the chain ladder method. Ensuring that all model assumptions are fulfilled we then compare estimates of reserves to estimates from the established chain ladder method. Finally we examine dependencies of lines of business within an insurance portfolio among the residuals.

This thesis is structured as follows. In Chapter 1 we give an introduction to reserving, the chain ladder method and the over-dispersed Poisson model. At the end of the chapter bootstrapping as a technique to simulate losses and estimate predictive reserve distributions within the chain ladder framwork is presented. In Chapter 2 we outline the theoretical framework of generalized additive models for location, scale and shape (GAMLSS). Chapter 3 gives an overview of distributions used in this thesis. In Chapter 4 we apply a GAMLSS to a line of business of an insurance portfolio and guide through the model fitting process. We show why the chain ladder method and generalized linear models fail in certain but common situations and why the GAMLSS is the better model in Chapter 5. We deal with the topic of prediction power in Chapter 6, where we compare next calendar year's cash flow projections between GAMLSS and chain ladder method. Finally, in Chapter 7, we analyze ultimate loss projections from GAMLSS and chain ladder method and examine residuals for dependence in an insurance portfolio.

1. Introduction to Reserving

In this chapter we present some basic definitions and results about loss triangles and reserving which are needed in the later chapters of this thesis. Main sources for this chapter are Mack (2002) and England and Verrall (2002) with proofs and more detailed explanations on the topics dealt with in this chapter. We start with a motivation for reserving and the structure of loss triangles (or run off triangles). We then introduce the chain ladder method to estimate reserves and a bootstrapping method based on a GLM with over-dispersed Poisson distribution.

1.1. The Necessity of Reserving

We start with a short motivation of reserving. An insurance company should be able to pay for claims whenever they occur. This is what clients expect when they sign a contract with the insurance company and what authorities verify throughout the year. On the other side, insurance companies probably don't need to put back enough money to pay for all claims of all clients at the same time, since the probability that this event occurs is vanishingly small. So it is the insurer's task to calculate the right amount of money to put back to pay those claims who are likely to occur. That has to be ensured on the short term as well as on the long term. This process is called reserving. The idea behind reserving is to estimate future obligations of the insurance company, mainly based upon their claims history. By that the insurance company gets an estimate of how much money they will have to pay in the next month, next year and eventually the next couple of decades. The money the insurance company has to put back then is called **reserve**. Estimation of reserves for each single contract would be tedious and difficult. On the other side, estimation for the whole portfolio at once would be too loose. By doing so one would assume that the future obligations of e.g. a motor own damage policy would behave similarly to the ones of e.g. a worker's compensation policy. Since this is not realistic, policies are aggregated into homogeneous groups, i.e. groups of policies with similar characteristics. Examples are 'all motor own damage policies' or 'all legal protection policies'. There exist different levels of aggregation and it is up to the actuary to choose an adequate level. Within this thesis we follow a relatively coarse grouping and divide the portfolio only into different lines of business, often called LoB's. Examples for lines of business are motor own damage, third party liability, legal protection, household contents insurances, etc. Reserving is then performed for each line of business and in most cases independently of other lines of business. A common and useful method for reserving is to analyze the claims history aligned in different types of triangles.

Accident Year i	Development Lag k 1	2	3	4	5
2007	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$
2008	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	
2009	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$		
2010	$x_{4,1}$	$x_{4,2}$			
2011	$x_{5,1}$				

Table 1.1.: General triangle structure

1.2. Loss Triangles

In order to estimate reserves insurance companies make use of their own claims history - if possible. The underlying assumption is that future claims will have a similar pattern as historical claims or that at least a functional relation between historical and future claims exists. A way of analyzing historical data is to align them in a triangle. Clearly, it is not the only way, but triangles have been used for a long time and proven to be a useful tool in order to estimate reserves. One (very small) example of such a triangle is shown in Table 1.1.

The rows of the triangle refer to the basis of calculation the data in the triangle is measured. In this example it is the **accident year**, i.e. the data in row '2010' belongs to the accident year 2010. A claim which occured in 2010 but has been reported in 2011 would then appear in the row for accident year 2010. To avoid long subscripts one may use accident years $1, \ldots, n$ instead, where 1 is the first available accident year (in this case 2007). Some other possibilities for the basis of calculation are the reporting year (i.e. in which year the claim has been reported to the company), the accounting year or the underwriting year (i.e. in which year the policy has been written). Depending on the underlying policies different calculation bases are useful. Here we will only deal with accident year based data.

The columns of the triangle refer to the different **development lags**. They indicate the time until the claim is known to the insurance company or a payment for the claim is made. E.g. a car accident at New Year's Eve 2010 happened in 2010 and hence appears in the row for accident year 2010. However, the insurance company won't be able to pay for that claim in 2010. If it paid on say 3.1.2011, the payment would technically have a lag. Development lags are measured in months, quarters or years and if not stated otherwise we have data on a yearly basis for development lags. A development lag of 1 year means that e.g. the claim is paid within the first year (i.e. the same year) after it occured. A lag of 2 years means the e.g. the claim is paid within the second year after the claims occured etc.. In Table 1.1 the cell in the last row of the first column $x_{5,1}$ belongs to accident year 2011, observed after the end of the first year, i.e. at the end of 2011.

We will denote the accident years by $i = 1, \ldots, n$ and the development lags by $k = 1, \ldots, n$. n is then the latest year for which data has been observed and is equal to the latest **calendar year** when using the accident year basis with development lags being one year.

Accident	Development Lag k			
Year i	1	2	3	4
1	2321	1142	-530	143
2	1645	889	745	
3	1868	996		
4	1574			

(a) Incremental losses

Accident	Development Lag k			
Year i	1	2	3	4
1	2321	3463	2933	3076
2	1645	2534	3279	
3	1868	2864		
4	1574			

(b) Cumulative losses

Table 1.2.: Example of triangles with incremental and cumulative losses

We therefore can write the triangle as the set $\{x_{i,k}|i + k \leq n + 1\}$.
Besides looking at a certain accident year (i.e. a certain row) or at a certain development lag (i.e. a certain column) it is sometimes useful to look at the diagonals of the triangles as well. In this context (and on a yearly basis) the diagonals are defined as sets $C_t = \{S_{i,k}|i + k - 1 = t\}$ for $t = 1, \dots, n$. Although being not real diagonals for $t < n$ the term 'diagonals' is commonly used in the insurance industry. Data in one diagonal belongs to the same calendar year. Looking at the calendar years is sometimes useful as some effects are only observable in this direction. An example could be a new law forcing the company to pay higher compensations to the clients. This would then be observable for all accident years and all development lags after the law has come into effect. Another example is inflation which can be an issue for long triangles, i.e. triangles with a long history.
We can now fill the triangle with data we want to analyze. Examples are incremental or cumulative payments, premiums, outstanding payments and many more. Again it depends on the line of business which triangles have to be analyzed. For the moment we only consider triangles with incremental or cumulative payments. The triangles are then called **loss triangles** and the payments are called **paid claims**.
Table 1.2 shows two triangles with incremental losses (a) and cumulative losses (b). For example, $x_{3,2} = 996$ in Table 1.2(a) is the sum of the payments made only at lag $k = 2$, i.e. in calendar year $t = i + k - 1 = 4$ for claims that originally occurred in accident year $i = 3$. Note that negative values for incremental payments are allowed and occur from time to time. However, the same cell $x_{3,2} = 2864$ in Table 1.2(b) is meant to be the sum of all payments until lag $k = 2$ for claims that originally occured in accident year $i = 3$.

Definition 1.1: *In general we denote* ***incremental payments*** *by $S_{i,k}$ and* ***cumulative payments*** *by $C_{i,k}$. Cumulative payments are defined as $C_{i,k} = \sum_{j=1}^{k} S_{i,j}$. The corresponding triangles are denoted by $\mathcal{S} = \{S_{i,k}|i = 1, \dots, n,\ k = 1, \dots, n + 1 - i\}$ and $\mathcal{C} = \{C_{i,k}|i = 1, \dots, n,\ k = 1, \dots, n + 1 - i\}$.*

Note: As mentioned above, negative values are allowed for incremental but not cumulative payments. In this context negative payments mean reimbursements to the insurance company. This can happen when e.g. it turns out that the insurance company is not liable to pay for a claim and gets a refund. Then for incremental paid losses at one development lag a huge refund could exceed the actual payments, leading to a negative incremental paid loss for that period. But clearly, the insurance company should not get back more money than they have paid, so cumulative paid losses should not be negative. ♢

1.3. Reserving

As outlined earlier, 'reserving' is the term used for 'estimating unknown future payments' and the reserve is the money put back by the insurance company to pay those estimated future obligations. Good estimation techniques for reserves are essential since

- too high reserves are unnecessarily bounded capital. That means the insurance company could have used a fraction of the capital for other investments with higher returns on capital.
- too low reserves can lead to serious problems for an insurance company. If not enough money has been reserved the insurance company might not be able to pay all obligations and could become insolvent.

It turned out that (loss) triangles are a very useful tool to estimate these payments and popular reserving methods like the chain ladder method make use of them. Having aligned data in a triangle structure like in Table 1.2, future payments are those payments located below the latest diagonal. Mainly they are caused by the fact that either

- a claim has occurred and has been reported to the company. The claim has been analyzed by the insurance company and the insurance company knows that it will have to pay money and how much they will have to pay. There is no uncertainty about the size of the payment but the payout may be deferred for different reasons. These claims are called then **outstanding claims**.
- a claim has already happened, but not been observed or reported to the company. An example for this are diseases caused by asbestos. Although the claim occurred a long time ago (i.e. when asbestos was used for building the house), diseases and claims occur many years later. If the insurance company supposes (by studying e.g. their claims history) that such losses can or will occur, they have to build reserves to pay these claims. The reserves are called **IBNR reserves**.

While uncertainty in the first case is zero it can be huge in the second case. A special case of the second are **IBNER reserves**. Assume a claim has occurred, been reported to the company, analyzed by the company and a reserve has been set up. However, unlike motor own damage claims the ultimate costs of e.g. a legal protection claim may differ materially from the initially estimated costs. The unsuccessful party could challenge the judgment and a new trial could cause much more costs to the insurance company than initially expected. If the insurance company fears that this could happen it has to build IBNER reserves, incurred but not enough reported reserves. This special case does not occur that often for most lines of business and hence is not analyzed separately from the IBNR analysis. It also depends on data availability and quality if a special IBNER analysis could be done at all. For the sake of simplicity we will only use IBNR reserves and assume IBNER reserves to be part of the IBNR reserves, which need no further analysis. Then the **ultimate loss** of a lines of business (or portfolio) is defined by

$$\text{Ultimate Loss} = \text{Paid Claims} + \text{Outstanding Claims} + \text{IBNR}. \tag{1.1}$$

Outstanding Claims and IBNR can be combined and just called reserve, so that

$$\text{Ultimate Loss} = \text{Paid Claims} + \text{Reserves}.$$

Definition 1.2: *Reserves for accident year i include all future obligations for this accident year and are calculated by*

$$R_i = \sum_{k=n+1-i}^{n} S_{i,k}, \qquad i = 2, \ldots, n.$$

R_i is call ***accident year reserve*** *for accident year i. The* ***total reserve*** *is the sum of all accident year reserves of the triangle, calculated by*

$$R = \sum_{i=2}^{n} R_i.$$

In order to run a business an insurance company must be able pay the ultimate loss. Since paid claims already have been managed (paid) in the past, focus lies on the reserves. We point out that reserves include future payments which may have to be made many years and decades after the initial estimation. How to ensure that payments can be made is an exercise for the financial department rather than for the actuarial department.

1.4. A Deterministic Approach: Chain Ladder Method

As times goes by, more and more claims are handled by the insurance company and all necessary payments are made. So eventually, when all claims are handled and no payments need to be done anymore, the paid claims equal the ultimate loss. The chain ladder method is based on this relationship. In actuarial literature it is often mentioned that the chain ladder method originated as a purely deterministic and computational algorithm. Though it leads to good estimates of future obligations in many cases and thus has been used for many years. Stochastic models for this algorithm have been developed later with Mack's distribution free model and his estimator for the standard error being the most important one. Hence we will first introduce the pure chain ladder method and afterwards derive Mack's stochastic model as well as a generalized linear model.

The main assumption is that future claims will behave similarly to past claims and hence the observed claims history can be used to estimate future claims. The method works on cumulative loss triangles $\mathcal{C}$ of the form like in Table 1.3.

Accident	Development Lag k				
Year i	1	2	$\cdots$	$n-1$	n
1	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	$\cdots$	$C_{2,n-1}$	
$\vdots$	$\vdots$	$\vdots$	$\iddots$		
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$			
n	$C_{n,1}$				

Table 1.3.: Cumulative paid claims triangle

At lag n all claims are assumed to be fully developed. That means no further claims occur after development lag n and all necessary payments have been made. For most lines of business and large triangles (i.e. with many years of claims history) this makes sense. An extension to this is to fit a tail curve which describes the development after the latest observable development lag, but this is out of the scope of this thesis and will not be handled. Thus, with the assumption that after n lags the claims are fully developed, the aim is to find estimates for ultimate losses $C_{i,n}$, $i = 2, \ldots, n$. $C_{1,n}$ does not need to be estimated and no reserve has to be set up for accident year $i = 1$. The reserves is then defined as

$$R_i = \sum_{k=n+2-i}^{n} S_{i,k} = C_{i,n} - C_{i,n+1-i} \quad i = 2, \ldots, n$$

and $R_1 = 0$ with $C_{i,n+1-i}$, $i = 2, \ldots, n$, being the latest known payment. To estimate $C_{i,n}$ different approaches exists. One them is the **multiplicative approach**, where the ultimate loss $C_{i,n}$ is modeled by

$$C_{i,n} = C_{i,1} \cdot F_{i,1} \cdot F_{1,2} \cdots F_{i,n-1},$$

where $F_{i,k} = C_{i,k+1}/C_{i,k}$ for all $k = 1, \ldots, n-1$ and $i = 1, \ldots, n$.

Notation: The factors

$$F_{i,k} = C_{i,k+1}/C_{i,k}, \quad k = 1, \ldots, n-1,\ i = 1, \ldots, n$$

are called **age-to-age factors**.

$F_{i,k}$ describes the loss development from development lag k to development lag $k+1$. Since we assume $C_{i,k} > 0$ for all $i, k = 1, \ldots, n$, this is well defined. $C_{i,k}$ is unknown for $i + k > n + 1$ and hence $F_{i,k}$ is unknown for $i + k > n + 1$.

To estimate unknown age-to-age factors the chain ladder method makes the assumption, that on average the age-to-age factors $F_{i,k}$ between accident years at one development lag are the same. The chain ladder method estimates theses average age-to-age factors, denoted by f_k, $k = 1, \ldots, n-1$, by volume weighted averages of observed age-to-age factors,

$$\hat{f}_k^{CL} := \frac{\sum_{i=1}^{n-k} C_{ik} F_{ik}}{\sum_{i=1}^{n-k} C_{ik}} = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{ik}}, \quad k = 1, \ldots, n-1, \tag{1.2}$$

where $C_{i,k}$ are the weights. These factors are called **chain ladder factors**. Then

$$\hat{C}_{i,n+2-i}^{CL} = C_{i,n+1-i} \cdot \hat{f}_{n+1-i}^{CL}, \quad i = 2, \ldots, n \tag{1.3}$$

and

$$\hat{C}_{i,k+1}^{CL} = \hat{C}_{i,k}^{CL} \cdot \hat{f}_k^{CL}, \quad i + k > n + 2 \tag{1.4}$$

are chain ladder estimates of unknown future cumulative payments. We can rewrite (1.4) as

$$\begin{aligned} \hat{C}_{i,k+1}^{CL} &= \hat{C}_{i,k}^{CL} \cdot \hat{f}_k^{CL} \\ &= \hat{C}_{i,k-1}^{CL} \cdot \hat{f}_{k-1}^{CL} \cdot \hat{f}_k^{CL} \\ &= \ldots = C_{i,n+1-i} \cdot \hat{f}_{n+i-1}^{CL} \cdots \hat{f}_k^{CL}, \quad i + k > n + 2. \end{aligned}$$

Hence ultimate losses can be estimated by

$$\hat{C}_{i,n}^{CL} = C_{i,n+1-i} \cdot \hat{f}_{n+1-i}^{CL} \cdots \hat{f}_{n-1}^{CL}, \quad i = 2, \ldots, n \tag{1.5}$$

and reserves by

$$\hat{R}_i^{CL} = \hat{C}_{i,n}^{CL} - C_{i,i+n-1} = C_{i,i+n-1} \cdot \left(\hat{f}_{i+n-1}^{CL} \cdots \hat{f}_{i,n-1}^{CL} - 1 \right), \quad i = 2, \ldots, n. \tag{1.6}$$

So with the assumption that on average the age-to-age factors at one development lag are the same for all accident year we get a relative easy estimator for the reserves. But a downside of this method clearly is that estimates for reserves yield no information about uncertainty.

1.5. Mack's Stochastic Model for the Chain Ladder Method

We stress at this point that the chain ladder factors are no result of a stochastic model but that stochastic models surrounding the chain ladder method have been developed after first appearance of the algorithm.

The chain ladder algorithm with estimators as defined in (1.2) does not account for any dependencies among accident years. That means the chain ladder method (implicitly) makes the following

Assumption (CL1): The chain ladder methods assumes that cumulative payments of different accident years are independent, i.e. that for $i, j = 1, \dots, n$,

$$\{C_{i,1}, \dots, C_{i,n}\}, \{C_{j,1}, \dots, C_{j,n}\}, i \neq j$$

are independent. $\diamondsuit$

From a statistical point of view it would have been preferable to have independently and identically distributed cumulative payments $C_{i,k}$. This assumption can be made, but it is not the *underlying* assumption of the chain ladder method. Also, the assumption of globally independent accident years is not necessarily fulfilled by the all triangles. Inflation or new laws can influence the triangle such that the assumption is violated. Hence a check to verify this assumption is inevitable.

Furthermore the chain ladder method assumes that on average the age-to-age factors at one development lag are the same for all accident years. If we consider age-to-age factors $F_{i,k}$ as random variables, this can be written as

$$\mathbb{E}\left[F_{i,k}\right] = f_k \quad k = 1, \dots, n-1. \tag{1.7}$$

As estimators for f_k, $k = 1, \dots, n-1$, we still use the chain ladder factors (1.2). So in this model we estimate unknown future payments again by

$$\hat{C}^{CL}_{i,k} = C_{i,n+1-i} \cdot \hat{f}^{CL}_{n+1-i} \cdots \hat{f}^{CL}_{k-1}, \quad i = 2, \dots, n,$$

and ultimate losses by

$$\hat{C}_{i,n} = C_{i,n+1-i} \cdot \hat{f}^{CL}_{n+1-i} \cdots \hat{f}^{CL}_{n-1}, \quad i = 2, \dots, n.$$

One can see that in this setup estimates of unknown future claims depend solely on the latest observable claim and the chain ladder factors. Even for the estimate of the ultimate loss no other (known) cumulative payments are taken into account. This can be written in terms of the conditional expectation:

$$\mathbb{E}\left[C_{i,k+1} | C_{i,1}, \dots, C_{i,k}\right] = C_{i,k} f_k \tag{1.8}$$

for $i = 2, \dots, n$, $k = 1, \dots, n-1$. Hence when using the chain ladder method we make (implicitly) the following

Assumption (CL2): There exist development factors $f_1, \dots, f_{n-1}$ such that

$$\mathbb{E}\left[C_{i,k+1}|C_{i1}, \dots, C_{i,k}\right] = C_{i,k} f_k \quad \forall i = 1, \dots, n,\ k = 1, \dots, n-1,$$

or equivalent to that

$$\mathbb{E}\left[\frac{C_{i,k+1}}{C_{i,k}}|C_{i1}, \dots, C_{i,k}\right] = f_k \quad \forall i = 1, \dots, n,\ k = 1, \dots, n-1,$$

The assumption is a special case of (1.7), compare Mack (2002)[p. 246]. The assumption CL2 leads to a case where we do not want to estimate $\mathbb{E}\left[C_{i,n}\right]$, $i = 2, \dots, n$ but $\mathbb{E}\left[C_{i,n}|\mathcal{C}\right]$, since

$$\begin{aligned}
\mathbb{E}\left[C_{i,n}|\mathcal{C}\right] &= \mathbb{E}\left[C_{i,n}|C_{i,k},\ i+k \leq n+1\right] \\
&\overset{CL1}{=} \mathbb{E}\left[C_{i,n}|C_{i,1}, \dots, C_{i,n+1-i}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[C_{i,n}|C_{i,1}, \dots, C_{i,n-1}\right]|C_{i,1}, \dots, C_{i,n+1-i}\right] \\
&\overset{CL2}{=} \mathbb{E}\left[C_{i,n-1} f_{n-1}|C_{i,1}, \dots, C_{i,n+1-i}\right] \\
&= \mathbb{E}\left[C_{i,n-1}|C_{i,1}, \dots, C_{i,n+1-i}\right] f_{n-1} \\
&= \dots \\
&= \mathbb{E}\left[C_{i,n+i-1}|C_{i,1}, \dots, C_{i,n+1-i}\right] f_{n+1-i} \cdots f_{n-1} \\
&= C_{i,n+i-1} \cdot f_{n+1-i} \cdots f_{n-1}.
\end{aligned}$$

Theorem 1.3: *Let*

$$\mathcal{C}_k := \{C_{i,j}|j \leq k,\ i+k \leq n+1\}, \quad 1 \leq k \leq n.$$

With assumptions CL1 and CL2 it holds that

(i) $\hat{f}_k^{CL}$ *is an unbiased estimator for* f_k *for all* $k = 1, \dots, n-1$

(ii) Conditioned on $\mathcal{C}_k$*, the* $\hat{f}_k^{CL}$*'s are uncorrelated with*

$$\mathbb{E}\left[\hat{f}_{n+1-i}^{CL} \cdots \hat{f}_{n-1}^{CL}|\mathcal{C}_{n+1-i}\right] = f_{n+1-i} \cdots f_{n-1}.$$

PROOF:

(i) Let

$$\mathcal{C}_k := \{C_{i,j} | j \leq k,\, i+k \leq n+1\}, \quad 1 \leq k \leq n$$

be the subset of the triangle $\mathcal{C}$ until development lag k, i.e. we cut off the triangle after development lag k. Then

$$\mathbb{E}\left[F_{i,k}|\mathcal{C}_k\right] \stackrel{CL1}{=} \mathbb{E}\left[F_{i,k}|C_{i,1},\ldots,C_{i,k}\right] \stackrel{CL2}{=} f_k.$$

Thus,

$$\begin{aligned}\mathbb{E}\left[\hat{f}_k^{CL}|\mathcal{C}_k\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n-k} C_{i,k}F_{i,k}}{\sum_{i=1}^{n-k} C_{i,k}}\middle|\mathcal{C}_k\right] = \frac{\sum_{i=1}^{n-k} C_{i,k}\mathbb{E}\left[F_{i,k}|\mathcal{C}_k\right]}{\sum_{i=1}^{n-k} C_{i,k}} \\ &= \frac{\sum_{i=1}^{n-k} C_{i,k}f_k}{\sum_{i=1}^{n-k} C_{i,k}} = f_k\end{aligned}$$

for all $k = 1,\ldots,n-1$. Hence

$$\mathbb{E}\left[\hat{f}_k^{CL}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{f}_k^{CL}|\mathcal{C}_k\right]\right] = \mathbb{E}\left[f_k\right] = f_k,$$

which shows that $\hat{f}_k^{CL}$ is an unbiased estimator for f_k.

(ii) To show that the $\hat{f}_k^{CL}$'s are conditional uncorrelated, let $j < k$. Then

$$\begin{aligned}\mathbb{E}\left[\hat{f}_j^{CL}\hat{f}_k^{CL}|\mathcal{C}_j\right] &= \mathbb{E}\left[\mathbb{E}\left[\hat{f}_j^{CL}\hat{f}_k^{CL}|\mathcal{C}_k\right]|\mathcal{C}_j\right] = \mathbb{E}\left[\hat{f}_j^{CL}\mathbb{E}\left[\hat{f}_k^{CL}|\mathcal{C}_k\right]|\mathcal{C}_j\right] \\ &\stackrel{(i)}{=} \mathbb{E}\left[\hat{f}_j^{CL}|\mathcal{C}_j\right] f_k \stackrel{(i)}{=} f_j f_k,\end{aligned}$$

so $\hat{f}_j^{CL}$ and $\hat{f}_k^{CL}$ are uncorrelated conditioned on $\mathcal{C}_j$. It then also follows that

$$\mathbb{E}\left[\hat{f}_{n+1-i}^{CL}\cdots\hat{f}_{n-1}^{CL}|\mathcal{C}_{n+1-i}\right] = f_{n+1-i}\cdots f_{n-1}.$$

$\square$

With the results of the theorem it can be shown that the chain ladder ultimate loss estimate (1.5) is an unbiased estimator for $\mathbb{E}\left[C_{i,n}|C_{i,1},\ldots,C_{i,n+1-i}\right]$ and the chain ladder reserve estimate (1.6) is an unbiased estimator for $\mathbb{E}\left[R_i|C_{i,1},\ldots,C_{i,n+1-i}\right]$, see Mack (1993) or Mack (2002).

The model defined so far surrounding the chain ladder method yields (condional) unbiased estimators of the reserves. Yet it does not include a measure of accucary of the estimate. Mack (1993) was able to give an explicit formula for the (conditional) mean squared error of the reserve estimate. Since then this measure has been used extensively. But to calculate the mean squared error, one more assumption has to be made. The conditional mean squared error for an estimate $\hat{R}_i$ of the unknown reserves R_i is defined as

$$\text{MSEC}\left[\hat{R}_i\right] = \mathbb{E}\left[\left(R_i - \hat{R}_i\right)^2 | C_{i,1}, \ldots, C_{i,n+1-i}\right].$$

This is not the same as the unconditional mean squared error

$$\text{MSE}\left[\hat{R}_i\right] = \mathbb{E}\left[\left(R_i - \hat{R}_i\right)^2\right].$$

We use the conditional mean squared error here because we want the best estimate of reserves for a given claims history, i.e. a given loss triangle. In contrast to that the unconditional mean squared error would average over all possible loss triangles.

Since $R_i = C_{i,n} - C_{i,n+1-i}$ and $\hat{R}_i = \hat{C}_{i,n} - C_{i,n+1-i}$,

$$\begin{aligned}
\text{MSEC}\left[\hat{R}_i\right] &= \mathbb{E}\left[\left(R_i - \hat{R}_i\right)^2 | C_{i,1}, \ldots, C_{i,n+1-i}\right] \\
&= \mathbb{E}\left[\left(C_{i,n} - \hat{C}_{i,n}\right)^2 | C_{i,1}, \ldots, C_{i,n+1-i}\right] \\
&= \text{MSEC}\left[\hat{C}_{i,n}\right].
\end{aligned}$$

To calculate the conditional mean squared error we need the following Proposition.

Proposition 1.4: *For random variables X, Y and a measurable function h it holds that*

$$\mathbb{E}\left[(X - h(Y))^2 | Y\right] = \mathbb{V}\text{ar}\left[X|Y\right] + \left(\mathbb{E}\left[X|Y\right] - h(Y)\right)^2.$$

With

$$\begin{aligned}
X &= C_{i,n} \\
Y &= (C_{i,1}, \ldots, C_{i,n+1-i}) \\
h(C_{i,1}, \ldots, C_{i,n+1-i}) &= \hat{C}_{i,n}
\end{aligned}$$

this becomes

$$\begin{aligned}
\text{MSEC}\left[\hat{C}_{i,n}\right] &= \mathbb{V}\text{ar}\left[C_{i,n} | C_{i,1}, \ldots, C_{i,n+1-i}\right] \\
&\quad + \left(\mathbb{E}\left[C_{i,n} | C_{i,1}, \ldots, C_{i,n+1-i}\right] - \hat{C}_{i,n}\right)^2.
\end{aligned} \tag{1.9}$$

From this equation it is obvious that

$$\hat{C}_{i,n} := \mathbb{E}\left[C_{i,n}|C_{i,1}, \ldots, C_{i,n+1-i}\right]$$

minimizes the second term and thus the conditional mean squared error. We stress that this form is true for $\hat{C}_{i,n}$ being deterministic. A slightly different version is shown in (1.13). But it also becomes clear the even if one knew the best estimator for $\hat{C}_{i,n}$, the conditional mean squared error still would have an unknown component $\mathbb{V}\mathrm{ar}\left[C_{i,n}|C_{i,1}, \ldots, C_{i,n+1-i}\right]$. So in order to calculate the conditional mean squared error the conditional variance is needed. Hence an assumption on the conditional variance is made. To motivate this, we use that with

$$G_{i,k} := \frac{C_{i,k}}{\sum\limits_{j=1}^{n-k} C_{j,k}}$$

the chain ladder factors can be written as

$$\hat{f}_k^{CL} = \frac{\sum\limits_{i=1}^{n-k} C_{i,k} F_{i,k}}{\sum\limits_{i=1}^{n-k} C_{i,k}} = \sum_{i=1}^{n-k} G_{i,k} F_{i,k}.$$

This is a weighted average of $F_{i,k}$'s with weights $G_{i,k}$. One can show (using Cauchy-Schwarz inequality) that weights are optimal in the sense that they minimize the conditional variance for $\hat{f}_k^{CL}$ if they are inversely proportional to $\mathbb{V}\mathrm{ar}\left[F_{i,k}|C_{i,k}\right]$. Since we assume that the chain ladder factors are optimal in the way we chose them, we implicitly make the following assumption.

Assumption (CL3): There exists constants $\sigma_1^2, \ldots, \sigma_{n-1}^2$ such that

$$\mathbb{V}\mathrm{ar}\left[\frac{C_{i,k+1}}{C_{i,k}}|C_{i,1}, \ldots, C_{i,k}\right] = \sigma_k^2 / C_{i,k}$$

or equivalently

$$\mathbb{V}\mathrm{ar}\left[C_{i,k+1}|C_{i,1}, \ldots, C_{i,k}\right] = C_{i,k}\sigma_k^2$$

holds for all $i = 1, \ldots, n$, $k = 1, \ldots, n-i$. $\diamondsuit$

With this last assumption the conditional mean squared error as defined in (1.9) can be calculated:

Theorem 1.5: *Under the assumptions CL1, CL2 and CL3 it holds that*

$$\mathrm{MSEC}(\hat{R}_i) = \hat{C}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum\limits_{j=1}^{n-k} C_{j,k}} \right).$$

Proof: See (Mack, 2002, Section 3.2.5). □

Analogously one can show a similar result for the total reserve $R = R_2 + \ldots R_n$:

Lemma 1.6: *Under the assumptions CL1, CL2 and CL3 it holds for* $\hat{R} = \hat{R}_2 + \ldots + \hat{R}_n$ *that*

$$\mathrm{MSEC}(\hat{R}) = \sum_{i=2}^{n} \left\{ \mathrm{MSEC}(\hat{R}_i) + \hat{C}_{i,n} \left(\sum_{k=i+1}^{n} \hat{C}_{k,n} \right) \sum_{k=n+1-i}^{n-1} \frac{2\hat{\sigma}_k^2}{\hat{f}_k^2} \frac{1}{\sum\limits_{l=1}^{n-k} C_{l,k}} \right\}.$$

Proof: See (Mack, 2002, Section 3.2.5). □

With these two results we extended the classic chain ladder method by a measure of accuracy of the estimates. We summarize the model with its assumptions in the following

Definition 1.7: *Let* $\mathcal{C} = \{C_{ik}, i+k \leq n+1\}$ *be a loss triangle with cumulative payments and let* $C_{ik} > 0$ *for all* i, k. *Let the following assumptions be fulfilled:*

(CL1) *The accident years are globally independent, i.e. for all* $i, j = 1, \ldots, n$, $i \neq j$,

$$\{C_{i,1}, \ldots, C_{i,n}\}, \{C_{j,1}, \ldots, C_{j,n}\}$$

are independent

(CL2) *There exist development factors* $f_1, \ldots, f_{n-1}$ *such that*

$$\mathbb{E}\left[C_{i,k+1} | C_{i1}, \ldots, C_{i,k}\right] = C_{i,k} f_k$$

holds for all $i = 1, \ldots, n$, $k = 1, \ldots, n-i$

(CL3) There exist constants $\sigma_1^2, \ldots, \sigma_{n-1}^2$ such that

$$\mathbb{V}\mathrm{ar}\left[C_{i,k+1} | C_{i,1}, \ldots, C_{i,k}\right] = C_{i,k}\sigma_k^2$$

holds for all $i = 1, \ldots, n$, $k = 1, \ldots, n-i$.

Then define

$$\hat{f}_k^{CL} := \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}, \quad k = 1, \ldots, n-1 \tag{1.10}$$

as the estimators for f_k. $\hat{f}_k^{CL}$ is called the ***chain ladder factor*** *for development lag k.*

Note (Connection to Weighted Least Squares Estimation):
(CL2) can interpreted as follows: fix development lag k, then (CL2) has the form of a linear regression of $C_{i,k+1}$ on $C_{i,k}$,

$$C_{i,k+1} = \alpha + \beta_k C_{i,k} + \varepsilon_i, \quad i = 1, \ldots, n-k,$$

with independent normal distributed error term ε_i with $\mathbb{E}\left[\varepsilon_i\right] = 0$. From (CL2) we get $\alpha = 0$ for all $i = 1, \ldots, n-k$ and from (CL3) we get $\mathbb{V}\mathrm{ar}\left[\varepsilon_i\right] = C_{i,k}\sigma_k^2$. That means variances are heteroscedastic, leading to weighted least squares estimation. Minimizing

$$Q(\beta_k) = \sum_{i=1}^{n-k} \frac{1}{C_{i,k}} \left(C_{i,k+1} - C_{i,k}\beta_k\right)^2$$

for $k = 1, \ldots, n-1$ indeed leads to the chain ladder factors as defined in (1.10), since

$$\begin{aligned} &\frac{\partial}{\partial \beta_k} Q(\beta_k) = \sum_{i=1}^{n-k} \frac{2}{C_{i,k}} \left(C_{i,k+1} - C_{i,k}\beta_k\right)\left(-C_{i,k}\right) \overset{!}{=} 0 \\ \Leftrightarrow\ &\frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}} = \beta_k. \end{aligned}$$

$\diamondsuit$

To summarize Mack's model, the main advantage is that this model does not require a specific distribution for incremental or cumulative payments. It makes only assumptions on the first two moments and one can quickly derive estimates of the reserves and the

conditional mean squared error of those estimates. It is important to keep in mind that all results have been derived conditioned on the loss triangle $\mathcal{C}$, especially the conditional mean squared error. In this model ultimate loss and reserve estimates are the same as for the original chain ladder method.
However, no loss or reserve distribution is available from this model. Other models, e.g. those within the framework of generalized linear models do not have this downside as we will show in the next section.

1.6. A Generalized Linear Model for the Chain Ladder Method

Mack's distribution free model is not the only stochastic model for the chain ladder method which yields the same estimates of reserves. Another popular approach is to fit the chain ladder method into the framework of generalized linear models.

Definition 1.8: *A* ***generalized linear model*** *(short: GLM) consists of*

Random Component

Random variables Y_i, $1 = 1, \ldots, n$, independently distributed with a probability density function or probability mass function coming from the exponential family with parameters θ and ϕ, given by

$$f_{Y_i}(y_i|\theta_i, \phi) = \exp\left\{\frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right\},$$

with known functions $a_i(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$. The Y_i's are called ***response variables****, θ_i the* ***canonical parameter*** *and ϕ the* ***dispersion parameter****.*

Systematic Component

A ***linear predictor***

$$\eta_i(\boldsymbol{\beta}) := \boldsymbol{x}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k}$$

with realizations $x_{i,j}$ of ***covariates*** *X_j, $j = 1, \ldots, k$, and unknown* ***regression parameter*** *$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^T$.*

Link Component

A known monotonic function $g : \mathbb{R} \to \mathbb{R}$ such that

$$g(\mu_i) = \eta_i(\boldsymbol{\beta}) = \boldsymbol{x}_i^T\boldsymbol{\beta},$$

where μ_i is the mean of Y_i. g is called ***link function****.*

In a more compact matrix notation the linear predictor can be written as

$$\boldsymbol{\eta}(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix} \boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta},$$

with $\boldsymbol{X} \in \mathbb{R}^{n \times k}$ the **design matrix**. The unknown regression parameter $\boldsymbol{\beta}$ has to be estimated from the data, e.g. by maximum likelihood estimation. It can be shown that the first two moments of an exponential family distribution are given by

$$\begin{aligned} \mathbb{E}\left[Y_i\right] &= b^{'}(\theta_i) =: \mu_i \\ \mathbb{V}\mathrm{ar}\left[Y_i\right] &= b^{''}(\theta_i) a_i(\phi) =: V(\mu_i) a_i(\phi), \end{aligned}$$

where $b^{'}(\theta_i) = \frac{\partial}{\partial \theta_i} b(\theta_i)$ and $b^{''}(\theta_i) = \frac{\partial^2}{\partial \theta_i^2} b(\theta_i)$. $V(\mu_i)$ is called the **variance function**. If $a_i(\phi)$ can be expressed as

$$a_i(\phi) = \frac{\phi}{\omega_i}$$

with *prior weights* ω_i, it is sometimes useful to express the variance by

$$\mathbb{V}\mathrm{ar}\left[Y_i\right] = \frac{\phi V(\mu_i)}{\omega_i}.$$

A special case of variance functions are power variance functions where

$$V(\mu_i) = \mu_i^{\xi},$$

with $\xi > 0$. Then the variance can be expressed as

$$\mathbb{V}\mathrm{ar}\left[Y_i\right] = \frac{\phi \mu_i^{\xi}}{\omega_i}.$$

In this specification quasi maximum likelihood estimation is performed to obtain parameter estimates. Special cases are e.g. $\xi = 0$, $\omega_i = 1$ for all i and $\phi > 0$, which relates to a GLM with Gaussian distribution and for $\xi = 1$, $\omega_i = 1$ for all i and $\phi = 1$, which describes a GLM with Poisson distribution. In these cases quasi maximum likelihood estimates are the same as maximum likelihood estimates.
A complete introduction and further results on generalized linear models can be found e.g. in McCullagh and Nelder (1989).

Based on Kremer (1982), Renshaw and Verrall (1994) defined a model which gives the same estimates of ultimate losses $C_{i,n}$, $i = 2, \ldots, n$, as the chain ladder method. Kremer suggested to use a log linear model for incremental payments $S_{i,k}$. In formulas,

$$\log(S_{i,k}) \sim \mathcal{N}\left(c + \alpha_i + \beta_k, \sigma^2\right) \qquad \text{independent},$$

with c, α_i, $\beta_k \in \mathbb{R}$ for $i = 1, \dots, n$, $k = 1, \dots, n$. α_i and β_k describe development among rows and columns, respectively, σ^2 is the variance. With

$$m_{i,k} := \mathbb{E}\left[\log(S_{i,k})\right], \qquad \phi = \sigma^2, \qquad \omega_{i,k} = \begin{cases} 1 & \text{, if } (i,k) \leq n+1 \\ 0 & \text{, else} \end{cases},$$

this can be interpreted as a GLM with linear predictor of the form

$$m_{i,k} = c + \alpha_i + \beta_k.$$

Instead of this Renshaw and Verrall (1994) used a generalized linear model with log link, but the same structure of the linear predictor, i.e.

$$\log(m_{i,k}) = \mu + \alpha_i + \beta_k,$$

where $m_{i,k} = \mathbb{E}\left[S_{i,k}\right]$. To specify the GLM, they set

$$V(\mu_i) := m_{i,k}, \qquad \omega_{i,k} := \begin{cases} 1 & \text{, if } (i,k) \leq n+1 \\ 0 & \text{, else} \end{cases}, \qquad \phi > 0.$$

By that interpretation of fitted values is easier than in Kremer's model since they already are on the normal scale and don't need to be transformed. Note that if ϕ is unknown this is not a GLM and quasi maximum likelihood estimation is required. Also, the model is over-parameterized and additional constraints are needed, e.g. demand $\alpha_1 = 0 = \beta_1$. Incremental payments can be estimated by

$$\hat{S}_{i,k} := \hat{m}_{i,k} = \exp\{\hat{c} + \hat{\alpha}_i + \hat{\beta}_k\},$$

with quasi maximum likelihood estimates $\hat{c}$, $\hat{\alpha}_i$, $\hat{\beta}_k$.
For $\phi > 1$ the model is an over-dispersed Poisson model. Recall that for a Poisson distribution it holds that

$$Y \sim Pois(\lambda) \quad \Rightarrow \quad \mathbb{E}\left[Y\right] = \lambda = \mathbb{V}\text{ar}\left[Y\right].$$

For the Poisson distribution over-dispersion means that the variance ist greater than the mean. So if Y has an over-dispersed Poisson distribution, it holds that

$$\mathbb{E}\left[Y\right] = \lambda, \quad \mathbb{V}\text{ar}\left[Y\right] = \phi\lambda,$$

with $\phi > 1$. Obviously, a Poisson distribution is not a useful distribution to model claim amounts. But since the solution of the quasi-likelihood equation does not rely on $C_{i,k}$'s which are non-negative integers, one does not care about the range of the underlying distribution in this special case, compare Mack and Venter (2000). So if we say that $S_{i,k}$ should follow an over-dispersed Poisson distribution, it holds that

$$\mathbb{E}\left[S_{i,k}\right] = m_{i,k} \quad \text{and} \quad \mathbb{V}\text{ar}\left[S_{i,k}\right] = \phi m_{i,k}$$

with $\phi > 1$. In this formulation the model just specifies the first two moments. To deal with over-dispersed Poisson models there are different ways. One possibility is to use quasi maximum likelihood estimation to estimate parameters. Another possibility is to use a so-called **mixing approach**, noted in McCullagh and Nelder (1989). For a random variable $S_{i,k}$ let

$$
\begin{aligned}
S_{i,k}|Z_{i,k} = z_{i,k} &\sim Pois(z_{i,k}) \\
Z_{i,k} > 0 \text{ with } \mathbb{E}\left[Z_{i,k}\right] &= m_{i,k} = \exp\{\alpha_i + \beta_k\}.
\end{aligned}
$$

The distribution of $Z_{i,k}$ is called the mixing distribution. Conditioned on an observation $z_{i,k}$, $Y_{i,k}$ then has a Poisson distribution with mean $z_{i,k}$. A popular distribution for $Z_{i,k}$ is the Gamma distribution with mean $m_{i,k}$ and index $\phi m_{i,k}$, i.e.

$$
\begin{aligned}
Z_{i,k} &\sim \Gamma(m_{i,k}, \phi m_{i,k}) \\
\mathbb{E}\left[Z_{i,k}\right] &= m_{i,k} \\
\mathbb{V}\text{ar}\left[Z_{i,k}\right] &= \frac{m_{i,k}}{\phi}.
\end{aligned}
$$

McCullagh and Nelder (1989) showed that in this case $S_{i,k}$ follows a negative binomial distribution with parameters $\phi m_{i,k}$ and $1/\phi$, i.e.

$$
\begin{aligned}
S_{i,k} &\sim negbin(\phi m_{i,k}, 1/\phi) \\
\mathbb{E}\left[S_{i,k}\right] &= m_{i,k} \\
\mathbb{V}\text{ar}\left[S_{i,k}\right] &= m_{i,k}(1 + \frac{1}{\phi}).
\end{aligned}
$$

We see that for every $\phi > 0$ over-dispersion is present and for $\phi \to \infty$ we observe a Poisson distribution.

Note: All of the following (theoretical) results are true for any model than can be used to describe an over-dispersed Poisson model. For the actual estimation we will use a GLM with negative binomal distribution. Thus, we use the term over-dispersed Poisson model and whenever estimations occur, this implicitly means they have been derived by a GLM with negative binomial distribution. Estimation in R can be done using `glm.nb` of the MASS package. $\diamondsuit$

Verrall (2000) showed that the reserve estimates for the over-dispersed Poisson model are the same as for the chain ladder method. We illustrate this in an example.

Example: Consider the following cumulative / incremental loss triangle in Table 1.4.

Accident Year	Development Lag 1	2	3	4	5
1	97	121	129	135	136
2	101	118	130	136	
3	100	122	130		
4	104	118			
5	101				

Accident Year	Development Lag 1	2	3	4	5
1	97	24	8	6	1
2	101	17	12	6	
3	100	22	8		
4	104	14			
5	101				

Table 1.4.: Cumulative and incremental loss triangle

We apply the chain ladder technique to the cumulative loss triangle and obtain Table 1.5.

Accident Year	Development Lag 1	2	3	4	5
1	97	121	129	135	136
2	101	118	130	136	**137.01**
3	100	122	130	**136.02**	**137.03**
4	104	118	**127.15**	**133.04**	**134.03**
5	101	**120.35**	**129.68**	**135.70**	**136.69**

Table 1.5.: Chain Ladder estimates (cumulative losses)

Then

$$\hat{R}_2 = 137.01 - 136 = 1.01, \quad \hat{R}_3 = 7.03, \quad \hat{R}_4 = 16.03, \quad \hat{R}_5 = 35.70$$

and hence

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \hat{R}_4 + \hat{R}_5 = 59.77.$$

For the over-dispersed Poisson model we can fit a generalized linear model with negative binomial distribution:

```
ex.nb <- glm.nb(formula=Incr~as.factor(DevLag) + as.factor(AYear),
                data=ex.data, link="log")
```

Note that this model uses incremental payments rather than cumulative payments, indicated by `Incr`. `DevLag` stands for the development lag, `AYear` for the accident year, both treated as factors. Predictions of the model are shown in Table 1.6.

Accident Year	Development Lag 1	2	3	4	5
1	97	24	8	6	1
2	101	17	12	6	**1.01**
3	100	22	8	**6.02**	**1.01**
4	104	14	**9.15**	**5.89**	**0.99**
5	101	**19.35**	**9.33**	**6.01**	**1.01**

Table 1.6.: GLM estimates (incremental losses)

It is easy to calculate that the estimated total reserve in this model is 59.77 and thus the same as for the chain ladder method. ♢

To get the mean squared error of $\hat{R}_i$ in this framework some more calculations have to be done. England and Verrall (2002) noted how to calculate the mean square error of $\hat{R}_i$ and $\hat{R}$ in this case. To ease notation, let

$$\Delta := \{(i,k)|i = 2,\ldots,n,\ k = n+2-i,\ldots,n\}$$

denote the indices of the unknown data below the latest diagonal, i.e. the lower triangle. Let

$$\Delta_i := \{(i,k)|k = n+2-i,\ldots,n\} \quad i = 2,\ldots,n$$

denote the i-row of the unknown triangle.

Proposition 1.10: *In the over-dispersed Poisson model, for $i = 2,\ldots,n$, the mean squared error of the estimate $\hat{R}_i$ of R_i is given by*

$$\begin{aligned} \mathrm{MSE}\left[\hat{R}_i\right] &= \sum_{k\in\Delta_i} \phi\hat{m}_{i,k} + \sum_{k\in\Delta_i} \hat{m}_{i,k}^2 \mathbb{V}\mathrm{ar}\left[\hat{\eta}_{i,k}\right] \\ &\quad + 2\sum_{\substack{k_1,k_2\in\Delta_i\\ k_1<k_2}} \hat{m}_{i,k_1}\hat{m}_{i,k_2}\mathrm{Cov}\left[\hat{\eta}_{i,k_1},\hat{\eta}_{i,k_2}\right]. \end{aligned} \tag{1.11}$$

The mean squared error of the estimate $\hat{R}$ of R is given by

$$\begin{aligned} \mathrm{MSE}\left[\hat{R}\right] &= \sum_{(i,k)\in\Delta} \phi\hat{m}_{i,k} + \sum_{(i,k)\in\Delta} \hat{m}_{i,k}^2 \mathbb{V}\mathrm{ar}\left[\hat{\eta}_{i,k}\right] \\ &\quad + 2\sum_{\substack{(i_1,k_1)\in\Delta\\ (i_2,k_2)\in\Delta\\ (i_1,k_1)\neq(i_2,k_2)}} \hat{m}_{i_1,k_1}\hat{m}_{i_2,k_2}\mathrm{Cov}\left[\hat{\eta}_{i_1,k_1}\hat{\eta}_{i_2,k_2}\right]. \end{aligned} \tag{1.12}$$

PROOF: See (England and Verrall, 2002, Section 7.2). □

It is important to note that this mean squared error is the unconditional mean squared error. This makes it very different from the conditional mean squared error that Mack proposed.

We see that for calculating the mean squared error also the variance-covariance matrix of $\hat{\eta}_{i,k}$, $(i,k) \in \Delta$, is required. This is not directly available but can be calculated from the variance-covariance matrix of the estimated parameters and the design matrix.

The over-dispersed Poisson model is one way to get a predictive distribution and mean squared error for ultimate losses or reserves. Other examples are the a negative binomial model or a normal approximation. In either case the models yield the same estimates of future claims. It is important to note that we use an *over-dispersed Poisson model*, which is only represented by a negative binomial distribution. This not the same as the *negative binomial model*, introduced in Verrall (2000).

1.7. Bootstrapping

As seen in previous sections the chain ladder method is an heuristic approach to determine reserves. Furthermore it provides measures for the variability of reserve estimate by the mean squared error (compare Theorem 1.5 and Lemma 1.6) when fit into Mack's stochastic model. But what it does not provide is a distribution of those estimates. The connection to generalized linear models in the last section guided a possible way to obtain a predictive loss distribution and especially a different way to estimate future claims. As for the chain ladder method the mean squared error can be calculated for these models, see Proposition 1.10. In general, for a random variable $S_{i,k}$ and it's estimate $\hat{S}_{i,k}$ the mean squared error of prediction can be written as

$$\mathrm{MSE}\left[\hat{S}_{i,k}\right] = \mathbb{E}\left[\left(S_{i,k} - \hat{S}_{i,k}\right)^2\right] = \mathbb{E}\left[\left(\left(S_{i,k} - \mathbb{E}\left[S_{i,k}\right]\right) - \left(\hat{S}_{i,k} - \mathbb{E}\left[S_{i,k}\right]\right)\right)^2\right].$$

The main difference to Proposition 1.4 is that here the estimate $\hat{S}_{i,k}$ is kept random and we do not consider a conditional mean squared error. Approximating $\mathbb{E}\left[S_{i,k}\right]$ by $\mathbb{E}\left[\hat{S}_{i,k}\right]$

in the second terms leads to an estimate of the mean squared error of the form

$$\begin{aligned}
\mathrm{MSE}\left[\hat{S}_{i,k}\right] &= \mathbb{E}\left[\left(S_{i,k} - \hat{S}_{i,k}\right)^2\right] \\
&\approx \mathbb{E}\left[\left(\left(S_{i,k} - \mathbb{E}\left[S_{i,k}\right]\right) - \left(\hat{S}_{i,k} - \mathbb{E}\left[\hat{S}_{i,k}\right]\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(S_{i,k} - \mathbb{E}\left[S_{i,k}\right]\right)^2\right] - 2\mathbb{E}\left[\left(S_{i,k} - \mathbb{E}\left[S_{i,k}\right]\right)\left(\hat{S}_{i,k} - \mathbb{E}\left[\hat{S}_{i,k}\right]\right)\right] \\
&\quad + \mathbb{E}\left[\left(\hat{S}_{i,k} - \mathbb{E}\left[\hat{S}_{i,k}\right]\right)^2\right] \\
&= \mathbb{V}\mathrm{ar}\left[S_{i,k}\right] - 2\mathrm{Cov}\left[S_{i,k}, \hat{S}_{i,k}\right] + \mathbb{V}\mathrm{ar}\left[\hat{S}_{i,k}\right].
\end{aligned}$$

If we assume that the future claim $S_{i,k}$ is independent of its estimator $\hat{S}_{i,k}$, the second term vanishes and we are left with

$$\begin{aligned}
\mathrm{MSE}\left[\hat{S}_{i,k}\right] &\approx \mathbb{V}\mathrm{ar}\left[S_{i,k}\right] + \mathbb{V}\mathrm{ar}\left[\hat{S}_{i,k}\right] \\
&\approx \text{process variance} + \text{estimation variance}.
\end{aligned} \tag{1.13}$$

Although calculation of the mean squared error is possible for the over-dispersed Poisson model (see Proposition 1.10), this guides an alternative way to obtain the mean squared error. If we estimated the process and estimation variance, the mean squared error could be estimated with (1.13). Furthermore, if we had the predictive distribution of the reserves, the mean squared error could be calculated as it's variance. Then also statistics like median and quantiles would be available and would allow for much more analyses of reserves.

But within e.g. the over-dispersed Poisson model the predictive distribution of accident year reserves $\hat{R}_i$, $i = 2, \dots, n$, is the sum of predictive distributions of the corresponding incremental payments. While for some cases the distribution of a sum of random variables can be easily calculated (e.g. if all random variables are Gaussian distributions), it is difficult to find the distribution in general in an analytical way.
But observe that the split into process variance and estimation variance of (1.13) is not present in expressions (1.11) and (1.12). So England and Verrall (1999) and England (2001) used the split into process variance and estimation variance and a bootstrapping technique to calculate the mean squared error from the sampled predictive values.

In general, bootstrapping means sampling with replacement from a given data set to obtain further information about statistics of interest. The statistics can be calculated for each of the sampled sets and a predictive distribution of the statistics can be estimated if the number of sampled sets is large enough.
In this context we resample the observed triangle and estimate reserves for each sampled triangle. Having done this often enough, a predictive distribution of the reserves can be estimated and the mean squared error as well as other statistics can be derived from the predictive distribution.

Note: In the following we are especially interested in calculating the mean squared error of an over-dispersed Poisson model. Hence we need to specify at this point what model we use. In our case, as mentioned earlier, we use a GLM with negative binomial distribution. To clearly differentiate our model from the negative binomial model introduced by Verrall (2000), we continue to call our model a GLM with negative binomial distribution rather than a negative binomial model. ♢

To apply bootstrapping, independent and identically distributed data is required. While we assume independence of the data for the GLM, data does not need to be identically distributed. Thus for bootstrapping with the GLM with negative binomial distribution we rather use scaled Pearson residuals r^{sP}, defined as

$$r_{i,k}^{sP} = \frac{S_{i,k} - \hat{m}_{i,k}}{\sqrt{\hat{\phi}\hat{m}_{i,k}}} \qquad i = 1, \dots, n, \; k = 1, \dots, n+1-i, \tag{1.14}$$

with $\hat{\phi}$ an estimate of the dispersion parameter ϕ, $S_{i,k}$ observed incremental payments and $\hat{m}_{i,k}$ estimated means. An estimator for ϕ is e.g.

$$\hat{\phi} = \frac{\sum_{i=1}^{n} \sum_{k=1}^{n+1-i} \left(r_{i,k}^{P}\right)^2}{\frac{1}{2}n(n+1) - 2n + 1}, \tag{1.15}$$

where

$$r_{i,k}^{P} = \frac{S_{i,k} - \hat{m}_{i,k}}{\sqrt{\hat{m}_{i,k}}} \qquad i = 1, \dots, n, \; k = 1, \dots, n+1-i,$$

are the unscaled Pearson residuals (see England and Verrall (2002)). In words, we divide the sum of unscaled Pearson residuals by the degrees of freedom, which is the number of observations ($\frac{1}{2}n(n+1)$) minus the number of parameters estimated ($2n-1$ = Intercept + $n-1$ factors for development lag + $n-1$ factors for accident year).
If the GLM is well specified the scaled Pearson residuals fulfill the requirements of independent and identically distributed data. Instead of one dispersion parameter for all observations more flexible approaches allowing for non-constant dispersion parameters can be used and may improve results. But then the model is no longer a GLM. Hence we use one parameter ϕ and preserve the GLM structure with the negative binomial distribution.

To estimate $\hat{m}_{i,k}$, $i = 1, \dots, n$, $k = 1, \dots, n+1-i$, two approaches exist:

- Use the fitted values of the GLM with negative binomial distribution
- Use a backward recursion as defined in (1.16)

As outlined in Renshaw and Verrall (1994), both methods yield the same estimate of $m_{i,k}$. While for the GLM statistical software is required, the backward recursion can be done in spreadsheets and is thus easier to handle from a practitioner's side. To apply the backward recursion, set

$$\begin{aligned}
\hat{C}_{i,n+1-i} &= C_{i,n+1-i} && i = 1, \ldots, n, \\
\hat{C}_{i,k} &= \hat{C}_{i,k+1} \cdot \frac{1}{\hat{f}_k^{CL}} && i = 1, \ldots, n-1, k = 1, \ldots, n-i \\
\hat{m}_{i,k} &= \hat{C}_{i,k} - \hat{C}_{i,k-1} && i = 1, \ldots, n-1, k = 2, \ldots, n+1-i \\
\hat{m}_{i,1} &= \hat{C}_{i,1} && i = 1, \ldots, n.
\end{aligned} \tag{1.16}$$

Since this may not be intuitive at first sight, we explore this in more detail in an example.

Example (Backward recursion): Consider a loss triangle with cumulative payments as in Table 1.7. If only incremental payments are available the cumulative payments can be calculated in the known way (compare Definition 1.1).

Accident Year i	Development Lag k 1	2	3
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$
2	$C_{2,1}$	$C_{2,2}$	
3	$C_{3,1}$		

Table 1.7.: 3x3 loss triangle with cumulative payments

Chain ladder factors are

$$\begin{aligned}
\hat{f}_1^{CL} &= \frac{C_{1,2} + C_{2,2}}{C_{1,1} + C_{2,1}} \\
\hat{f}_2^{CL} &= \frac{C_{1,3}}{C_{1,2}}.
\end{aligned}$$

We apply the backward recursion:

Step 1: Start with latest diagonal

Accident Year i	Development Lag k 1	2	3
1			$\hat{C}_{1,3} = C_{1,3}$
2		$\hat{C}_{2,2} = C_{2,2}$	
3	$\hat{C}_{3,1} = C_{3,1}$		

Table 1.8.: Step 1

Step 2: Recursively go backwards...

Accident Year i	Development Lag k 1	2	3
1		$\hat{C}_{1,2} = \hat{C}_{1,3}/\hat{f}_2^{CL}$	$\hat{C}_{1,3}$
2		$\hat{C}_{2,2}$	
3	$\hat{C}_{3,1}$		

Table 1.9.: Step 2

Step 3: ...until first column is filled

Accident Year i	Development Lag k 1	2	3
1	$\hat{C}_{1,1} = \hat{C}_{1,2}/\hat{f}_1^{CL}$	$\hat{C}_{1,2}$	$\hat{C}_{1,3}$
2	$\hat{C}_{2,1} = \hat{C}_{2,2}/\hat{f}_1^{CL}$	$\hat{C}_{2,2}$	
3	$\hat{C}_{3,1}$		

Table 1.10.: Step 3

Step 4: Calculate $\hat{m}_{i,k}$ by differencing.

Accident Year i	Development Lag k 1	2	3
1	$\hat{m}_{1,1} = \hat{C}_{1,1}$	$\hat{m}_{1,2} = \hat{C}_{1,2} - \hat{C}_{1,1}$	$\hat{m}_{1,3} = \hat{C}_{1,3} - \hat{C}_{1,2}$
2	$\hat{m}_{2,1} = \hat{C}_{2,1}$	$\hat{m}_{2,2} = \hat{C}_{2,2} - \hat{C}_{2,1}$	
3	$\hat{m}_{3,1} = \hat{C}_{3,1}$		

Table 1.11.: Step 4

$\diamondsuit$

The procedure so far has been the following:

- input: a loss triangle with incremental payments
- estimate the GLM with a negative binomial distribution
- calculate scaled Pearson residuals $r^{sP}_{i,k}$ and $\hat{m}_{i,k}$ for $i = 1, \dots, n,\ k = 1, \dots, n+1-i$

As mentioned above the reason for using this recursion is the fact that it can be easily done in a spreadsheet. No matter which of both approaches is used to calculate $\hat{m}_{i,k}$ for $i = 1, \dots, n,\ k = 1, \dots, n+1-i$, we are still interested in incremental payments. From (1.14) it follows that

$$S_{i,k} = r^{sP}_{i,k} \sqrt{\hat{m}_{i,k}} + \hat{m}_{i,k} \qquad i = 1, \dots, n,\ k = 1, \dots, n+1-i. \tag{1.17}$$

Resampling the scaled Pearson residuals leads to a new set of incremental payments using (1.17). To identify the resampled data in each iteration, let B denote the number of total iterations done in the bootstrap, e.g. B=1000. Let $b = 1, \dots, B$ be the current iteration. Then a new incremental payment of accident year i at lag k in iteration b is given by

$$S^b_{i,k} = \left(r^{sP}_{i,k}\right)^b \sqrt{\hat{m}^b_{i,k}} + \hat{m}^b_{i,k} \qquad i = 1, \dots, n,\ k = 1, \dots, n+1-i.$$

Let

$$\mathcal{S}^b = \{S^b_{i,k} | i = 1, \dots, n,\ k = 1, \dots, n+1-i\}$$

denote the triangle of resampled incremental payments in iteration b and

$$\mathcal{C}^b = \{C^b_{i,k} | i = 1, \dots, n,\ k = 1, \dots, n+1-i\}$$

the corresponding triangle of cumulative payments.
From each triangle $\mathcal{S}^b$ ($\mathcal{C}^b$) unknown payments $\hat{S}^b_{i,k}$ ($\hat{C}^b_{i,k}$) in the lower triangle can be estimated; the GLM uses $\mathcal{S}^b$, the chain ladder method $\mathcal{C}^b$. Then, accident year reserves

$\hat{R}_i^b$, $i = 2, \ldots, n$ and total reserve $\hat{R}^b$ can be estimated. Again both methods lead to the same result but the chain ladder method is faster and estimation can be done in a spreadsheet.
Having a set of total reserves $\{\hat{R}^1, \ldots, \hat{R}^B\}$, the mean of this set could be calculated and used as the reserve estimate of the triangle. But this does not yet specify a predictive distribution of the reserve and hence we can not calculate the mean squared error. Recalling (1.13), the sample variance of $\{\hat{R}^1, \ldots, \hat{R}^B\}$ is the estimation variance. We still need to calculate process variance. Since this is not observable, a second stage of the bootstrap is needed.

In each iteration b and for every $\hat{S}_{i,k}^b$ in the lower triangle we simulate once from the underlying distribution; in this case negative binomial distribution with mean $m_{i,k}^b$. $m_{i,k}^b$ is unknown and needs to estimated from the data. As for estimates $\hat{m}_{i,k}$ of $m_{i,k}$ in the upper triangle, there are two ways of estimating $m_{i,k}^b$ in the lower triangle:

- by the fitted values of the over-dispersed Poisson model (GLM)
- a backward recursion as defined in (1.18).

In the second case, we set for $i = 2, \ldots, n$, $k = n + 2 - i, \ldots, n$

$$\begin{aligned} \hat{C}_{i,k+1}^b &= \hat{C}_{ik}^b \left(\hat{f}_k^{CL}\right)^b \\ \hat{m}_{ik}^b &= \hat{C}_{i,k}^b - \hat{C}_{i,k-1}^b. \end{aligned} \tag{1.18}$$

The first line of (1.18) is nothing else but the chain ladder method (with chain ladder factors calculated from the resampled triangle) and the second one is simple differencing to obtain incremental payments and hence estimates for $m_{i,k}^b$. The recursion can obviously be done in a spreadsheet and might be easier to use. But again both methods yield the same estimate.

The estimates $\hat{m}_{i,k}^b$ refer to the means of the negative binomial distribution. So the unknown future payments including the process variance, denoted by $\tilde{S}_{i,k}^b$, can then be estimated by simulating once from this distribution:

$$\tilde{S}_{i,k}^b \sim negbin(\hat{\phi}\hat{m}_{i,k}^b, 1/\hat{\phi}).$$

Having a set $\{\tilde{S}_{i,k}^1, \ldots, \tilde{S}_{i,k}^B\}$ of such payments a predictive distribution could be estimated. But since we are interested in the predictive distribution of reserves, let

$$\tilde{R}_i^b := \sum_{k=n+2-i}^{n} \tilde{S}_{i,k}^b$$

and

$$\tilde{R}^b := \sum_{i=2}^{n} \tilde{R}_i^b$$

be the (accident year) reserves at iteration b. From the sets $\{\tilde{R}_i^1, \ldots, \tilde{R}_i^B\}$ or $\{\tilde{R}^1, \ldots, \tilde{R}^B\}$ we can estimate the predictive distribution and can calculate mean, variance and confidence intervals. E.g. for $\tilde{R}_i$, the sample mean

$$\tilde{R}_i^{BE} := \frac{1}{B}\sum_{b=1}^{B} \tilde{R}_i^b$$

could be used as a best estimate for R_i and could be compared to the chain ladder estimate. The sample variance is an estimator for the mean squared error and could then be compared to the mean squared error from the Mack's stochastic model for the chain ladder method. But again one has to keep in mind that this mean squared error derived by bootstrapping is an unconditional mean squared error whereas Mack's mean squared error is a conditional one. We summarize the procedure at the end of this section in Algorithm 1.1.

The output of the bootstrap as given in Algorithm 1.1 are sets of estimated accident year reserves. To get sets of total reserves another summation in line 15 by accident year is required. As mentioned above results could be compared to the chain ladder method. However, since we apply another statistical model to loss triangles in the next chapters, we can not only compare the first two moments but the complete predictive distribution of both models.

For more about the bootstrap for the over-dispersed Poisson model as well as for other models we refer to England and Verrall (2002). An implementation of the chain ladder method with bootstrapping for R is provided by Markus Gesmann, Daniel Murphy and Wayne Zhang in their `ChainLadder` package (see `http://code.google.com/p/chainladder/`).

Algorithm 1.1 Bootstrap with B simulations for ODP model

Require: Loss Triangle with cumulative payments
Ensure: Sets of reserve estimates by accident year

Calculate chain ladder factors $\hat{f}_k^{CL}$ from cumulative loss triangle by

$$\hat{f}_k^{CL} = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{ik}}, \quad k = 1, \ldots, n-1.$$

Calculate unscaled Pearson residuals of the upper triangle by

$$r_{i,k}^P = \frac{S_{i,k} - \hat{m}_{i,k}}{\sqrt{\hat{m}_{i,k}}}$$

Calculate $\hat{\phi}$ as in (1.14)
Calculate scaled Pearson residuals in the upper triangle by (1.15)
Calculate $\hat{m}_{i,k}$ in the upper triangle using backward recursion and differencing as defined in (1.16)
Begin iterative loop
for $l = 1$ to B **do**
  resample scaled Pearson residuals $\left(r_{i,k}^{sP}\right)^b$
  determine new incremental payments $S_{i,k}^b$, $i = 1, \ldots, n$, $k = 1, \ldots, n+1-i$ by

$$S_{i,k}^b = \left(r_{i,k}^{sP}\right)^b \sqrt{\hat{m}_{i,k}^b} + \hat{m}_{i,k}^b \qquad i = 1, \ldots, n,\ k = 1, \ldots, n+1-i.$$

  determine corresponding incremental and cumulative loss triangles $\mathcal{S}^b$ and $\mathcal{C}^b$
  calculate chain ladder factors $\left(\hat{f}_k^{CL}\right)^b$
  calculate cumulative payments $\hat{C}_{i,k}^b$ in the lower triangle with the chain ladder factors.
  calculate $\hat{m}_{ik}^b$ for in the lower triangle using

$$\hat{m}_{ik}^b = \hat{C}_{i,k}^b - \hat{C}_{i,k-1}^b$$

  for each cell in the lower triangle simulate once from an over-dispersed Poisson distribution with mean $\hat{m}_{i,k}^b$ as calculated in the last step to get $\tilde{S}_{i,k}^b$
  calculate accident year reserve $\tilde{R}_i^b$ by

$$\tilde{R}_i^b := \tilde{S}_{i,n+2-i}^b + \ldots + \tilde{S}_{i,n}^b$$

  store $\{\tilde{R}_i^b, \ldots, \tilde{R}_n^b\}$
end for

2. Introduction to GAMLSS

In this section we introduce the framework of generalized additive models for location, scale and shape, GAMLSS. We will focus on the theoretical part of it in this section.

2.1. The Idea

GAMLSS were introduced by Rigby and Stasinopoulos (2001, 2005). They have been established to allow for more flexibility when modeling data of all kind than a generalized linear model. For generalized linear models (GLM), compare Definition 1.7,

- the distribution of the response variable Y has to come from an exponential family. The probability density function (pdf) or probability mass function (pmf) of Y can be written as

$$f_Y(y|\theta, \phi) = \exp\left\{\frac{\theta y - b(y)}{a(\phi)} + c(y, \phi)\right\},$$

  where ϕ is the **dispersion paramater**, θ the **canonical parameter** and $a(\cdot)$, $b(\cdot)$ and $c(\cdot, \cdot)$ are known functions.

- there is a **linear predictor** of the form

$$\eta = \eta(\boldsymbol{\beta}) = \boldsymbol{x}^T\boldsymbol{\beta} = \beta_0 + x_1\beta_1 + \ldots + x_k\beta_k$$

  with covariates $x_1, \ldots, x_k$ and unknown regression parameters $\beta_0, \ldots, \beta_k$.

- the relation between the mean $\mu = \mathbb{E}[Y]$ and the linear predictor η is given by a **link function** g such that

$$g(\mu) = \eta = \boldsymbol{x}^T\boldsymbol{\beta}.$$

An extension of GLM are generalized additive models (GAM). They are more flexible by allowing the (transformed) mean $g(\mu)$ not only to be a linear combination of covariates $x_1, \ldots, x_k$, but to be a sum of smooth functions of covariates. That is we allow for smooth functions $f_1, \ldots, f_k$ such that

$$g(\mu) = \beta_0 + f_1(x_1) + \ldots + f_k(x_k),$$

where $f_1, \ldots, f_k$ can have a parametric or non-parametric form. We refer to Hastie and Tibshirani (1990) for a detailed discussion of GAM.

Although GLM and GAM are extensive models which may work well in many cases, there are cases where even more flexible models are desirable. For a GLM the first two central moments of the exponential family are given as

$$\mathbb{E}\left[Y\right] = b'(\theta) \quad \text{and} \quad \mathbb{V}\text{ar}\left[Y\right] = b''(\theta)a(\phi),$$

where $b'(\theta)$ and $b''(\theta)$ are the first and second derivate of $b(\theta)$ w.r.t. to θ, respectively. $b''(\theta)$ is called **variance function**. We can see that the variance is a function of the mean, defined implicitly via the variance function. Rigby and Stasinopoulos (2005) claim that this is true for skewness and kurtosis as well. Thus the idea has been to develop a new model which allows explicit modeling of these moments rather than keeping implicite dependence on the mean. They also relaxed the requirement of a distribution from an exponential family by allowing more general distributions.

2.2. The Model

Let n be the number of observations of a random variable Y. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ be the corresponding random vector. We assume that Y_i, $i = 1, \ldots, n$, are independent distributed conditioned on a vector $\boldsymbol{\theta}^{i^T} = (\theta_{i1}, \ldots, \theta_{ip})$ of p parameters and have pdf or pmf $f_{Y_i}(y_i|\boldsymbol{\theta}^i)$. $\boldsymbol{\theta}^{i^T}$, $i = 1, \ldots, n$, are sets of the overall population parameter vector $\boldsymbol{\theta}^T = (\theta_1, \ldots, \theta_p)$ related to the ith observation. In this thesis we deal with a maximum of four parameters $\theta_1, \ldots, \theta_4$ and call them μ, σ, ν and τ, respectively. μ is meant to be the location, σ the scale and ν and τ the shape parameters. In most cases they are closely related to the first four (central) moments. As an example consider a Gaussian distribution. Naturally there a two parameters, μ the mean and σ the standard deviation. ν and τ aren't necessary at all such that $p = 2$.
As for GLM and GAM we need to define a functional relation between the parameters and covariates. In general, let $g_1(\cdot), \ldots, g_4(\cdot)$ be known monotonic link functions such that

$$\begin{aligned}
g_1(\boldsymbol{\theta}_1) &= g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \boldsymbol{Z}_{j1}\gamma_{j1} \\
g_2(\boldsymbol{\theta}_2) &= g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \boldsymbol{Z}_{j2}\gamma_{j2} \\
g_3(\boldsymbol{\theta}_3) &= g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \boldsymbol{Z}_{j3}\gamma_{j3} \\
g_4(\boldsymbol{\theta}_4) &= g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \boldsymbol{Z}_{j4}\gamma_{j4},
\end{aligned} \tag{2.1}$$

where for $k = 1, 2, 3, 4$, $\boldsymbol{\theta}_k^T = (\theta_{1k}, \ldots, \theta_{nk})$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$ are vectors of length n, $\boldsymbol{X}_k \in \mathbb{R}^{n \times J'_k}$ are known design matrices (for the fixed effects), $\boldsymbol{\beta}_k^T = (\beta_{1k}, \dots, \beta_{J'_k})$ are parameter vectors (for the fixed effects), $\boldsymbol{Z}_{jk} \in \mathbb{R}^{n \times q_{jk}}$ are known design matrices (for the random effects) and γ_{jk} are random vectors of dimension q_{jk}. We then call this model a **Generalized Additive Model for Location, Scale and Shape (GAMLSS)**. Because of numerical advantages (compare Rigby and Stasinopoulos (2005, Appendix B)), the vectors γ_{jk}, $j = 1, \dots, J_k$, are not combined into a single vector γ_k. Each linear predictor $\boldsymbol{\eta}_i$ consists of

- a parametric component $\boldsymbol{X}_i\boldsymbol{\beta}_i$
- an additive (random) component $\sum_{j=1}^{J_i} \boldsymbol{Z}_{ji}\gamma_{ji}$.

Instead of random effects γ_{ji} one could also use smooth functions like for GAM. In the R package `gamlss` there are currently many kinds of additive terms implemented. To name some of them cubic splines, penalized splines, varying coefficients, LOESS and random effects are available and offer a maximum degree of flexibility. Each $\boldsymbol{\eta}_i$ can be modeled using its own specific set of covariates, additive terms and link function. This allows modeling more complex scenarios than e.g. GLM or GAM.
In this thesis we will use a slightly simpler version of model (2.1) by not using random effects. That is we will use the model

$$\begin{aligned} g_1(\boldsymbol{\theta}_1) &= g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 \\ g_2(\boldsymbol{\theta}_2) &= g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 \\ g_3(\boldsymbol{\theta}_3) &= g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 \\ g_4(\boldsymbol{\theta}_4) &= g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4, \end{aligned} \tag{2.2}$$

and refer to this as the (fully parametric) GAMLSS. It is still more flexible than a GLM or GAM because although loosing the possibility of having additive terms we preserve the opportunity to model up to four parameters and use an almost arbitrary distribution for $\boldsymbol{Y}$. The only requirement for the distribution is that the (log-)likelihood and the first and second derivatives have to be available and computable since maximum likelihood estimation is used to fit a model. Since we not allow for more than four parameters the distribution should also not have more than four parameters.

Notation: To identify a GAMLSS for a random variable Y we use the notation in line with Rigby and Stasinopoulos (2005),

$$Y \sim \mathcal{D}\left(g_1(\theta_1) = t_1,\, g_2(\theta_2) = t_2,\, g_3(\theta_3) = t_3,\, g_4(\theta_4) = t_4\right),$$

where

- $\mathcal{D}$ defines the distribution of Y. We will only use distributions we introduce in Chapter 3.
- $g_1(\cdot), \dots, g_4(\cdot)$ are the known monotonic link functions for parameters $\theta_1, \dots, \theta_4$.

- $t_1, \ldots, t_4$ are the model formulae for the explanatory terms. As mentioned in above, we will only use a parametric component without additive terms, see (2.2).

Consider the following example:

$$Y \sim \mathrm{TF}\left(\mu = poly(X_1, 2),\, \ln(\sigma) = X_2,\, \nu = 1\right).$$

This would be a model for Y where Y is assumed to follow a t-distribution. The location parameter μ is modeled by a polynomial on covariate X_1 of degree 2 with identity link, the scale parameter σ is modeled by covariate X_2 with a log-link and the parameter ν which describes the degree of freedom of the t-distribution is modeled as a constant (with identity link).
Note that in this formulation $\nu = 1$ does not mean to fix ν at 1 but to estimate a constant for ν. To fix a certain paramater we denote this e.g. by

$$Y \sim \mathrm{TF}\left(\ln(\mu) = X_1 + as.factor(X_2),\, \ln(\sigma) = poly(X_2, 3),\, \nu := 2\right).$$

In this model Y has again t-distribution, where μ is modeled by X_1 and a factor on X_2 with log link, σ by a polynomial on X_2 of degree 3 with log link and ν is fixed at 2.

At this point we end the brief introduction of GAMLSS. A detailed discussion can be found in Rigby and Stasinopoulos (2005), also making notes about the distributions, algorithms and model selection process. The choice of the distribution depends strongly on the data and for our insurance portfolio some non-standard distributions have been used. Therefore we introduce the set of distributions used in this thesis in Chapter 3.

3. Distributions

In this section, we want to give a brief overview about the distributions used for fitting the Paid-to-Premium (PtP) ratios. We only state results about these partially not very common distributions and/or parameterizations. More discussion about the distributions can be found in the literature.

3.1. Gaussian Distribution

For the sake of completeness we start with the Gaussian distribution (and the t-distribution afterward)

Definition 3.1: *A random variable Y has a **Gaussian distribution** if its probability density function has the form*

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}, \tag{3.1}$$

where $y \in \mathbb{R}$, $\mu \in \mathbb{R}$ is the mean and $\sigma > 0$ is the standard deviation (or volatility). We then write $X \sim \mathcal{N}(\mu,\sigma)$.

Note that a slightly different parametrization is given by

$$f_Y(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right\} \tag{3.2}$$

where $y \in \mathbb{R}$, $\mu \in \mathbb{R}$ is the mean, $\sigma^2 > 0$ is the variance and we then write $X \sim \mathcal{N}(\mu,\sigma^2)$. This has to be mentioned since both version are implemented in the gamlss package. The first one corresponds to the `NO`-distribution and will be used throughout this thesis. The second corresponds to the `NO2`-distribution.
Using the first version (3.1) the mean and variance are

$$\begin{aligned}\mathbb{E}\left[Y\right] &= \mu\\ \mathbb{V}\mathrm{ar}\left[Y\right] &= \sigma^2,\end{aligned}$$

which exist for all valid choices of μ and σ.

3.2. t-Distribution

We need to introduce two version of the t-distribution in this section: the standardized and the non-standardized t-distribution.

Definition 3.2: *A random variable Y has a **(standardized) t-distribution** if its probability density function is*

$$f_Y(y|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{3.3}$$

for $y \in \mathbb{R}$, where $\nu > 0$ is the number of the degrees of freedom. Equivalently one could write

$$f_Y(y|\nu) = \frac{1}{\sqrt{\nu}B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{3.4}$$

where $B(a,b)$ is the Beta function. We then write $X \sim t(\nu)$.

Mean and variance are given by

$$\begin{aligned} \mathbb{E}\left[Y\right] &= 0 && \text{if } \nu > 1 \\ \mathbb{V}\text{ar}\left[Y\right] &= \frac{\nu}{\nu-2} && \text{if } \nu > 2. \end{aligned}$$

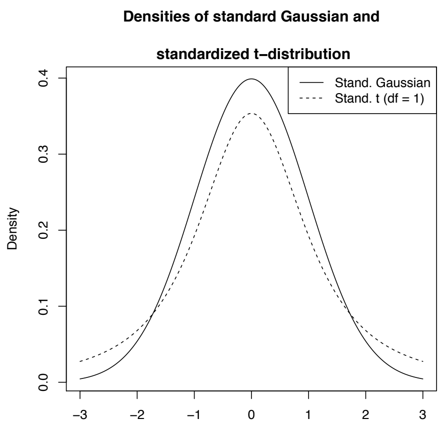


Figure 3.1.: Density plot for standard Gaussian and standardized t-distribution

A plot of densities for the standard Gaussian and standardized t-distribution is given in Figure 3.1. The t-distribution has heavier tails than the Gaussian distribution for small degrees of freedom. For increasing degrees of freedom the standardized t-distribution

approaches the standard Gaussian distribution. While for 30 degrees of freedom an approximation by the standard Gaussian distribution is enough for most cases the gamlss package uses the standard Gaussian distribution not until 10^6 degrees of freedom.
A generalization of (3.3) is the non-standardized t-distribution.

Definition 3.3: *A random variable Y with probability density function*

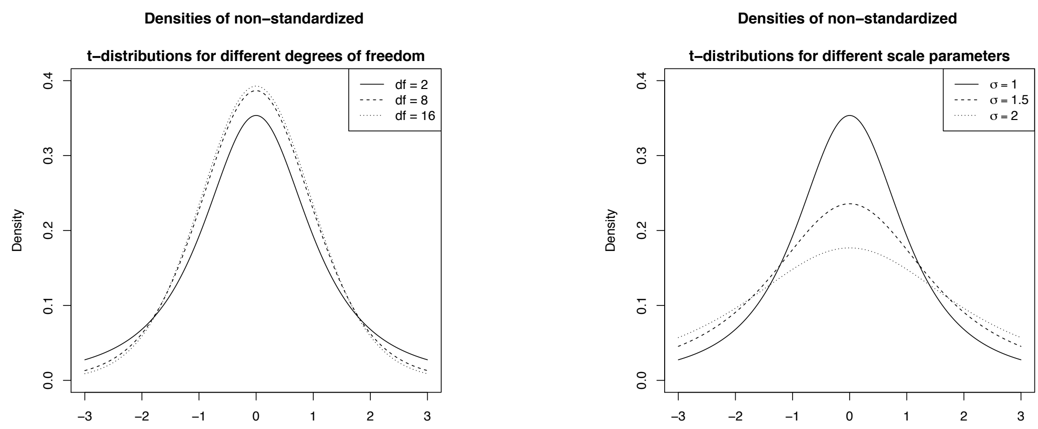
$$f_Y(y|\mu,\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{3.5}$$

*where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\nu > 0$ the number of degrees of freedom is called **non-standardized t-distribution** . We then write $Y \sim t(\mu,\sigma,\nu)$.*

Mean and variance are given by

$$\begin{aligned} \mathbb{E}\left[Y\right] &= \mu & \text{if } \nu > 1 \\ \mathbb{V}\text{ar}\left[Y\right] &= \sigma^2\frac{\nu}{\nu-2} & \text{if } \nu > 2. \end{aligned}$$

Jackman (2009) noted some more details about this distribution. In the gamlss package both distributions are included in the `TF`-distribution, with the standardized t-distribution being a special case of the non-standardized t-distribution with $\mu = 0$ and $\sigma = 1$.



(a) A small number of degrees of freedom leads to heavy tails

(b) The variance increases with increasing σ and the curve flatens

Figure 3.2.: Density plots of non-standardized t-distributions with different degrees of freedom ν or scale parameter σ

Figure 3.2 shows that tails are heavier the fewer the degrees of freedom are. Since we are interested in modeling the scale parameter with a GAMLSS we require at least two degrees of freedom to guarantee existence of the variance. For increasing scale parameter we see a flattening density curve, analog to the Gaussian case.

3.3. Zero-Inflated Distributions

Almost all analyses of loss triangles known to us were performed on triangles with 'good' losses. That means neither losses of size 0 nor negative losses occurred. But since both cases occur (often) in reality we need to model them. The second case of negative losses can easily be included by using e.g. the above introduced Gaussian and t-distributions. The first case is not that easy but can be modeled by using a mixed discrete-continuous distribution. We will introduce two of them now.
A zero-inflated distribution is mixture of a continuous distribution with a discrete distribution. In this work Gaussian or t-distributions will play the role of the continuous distributions. The discrete distribution is of type Bernoulli. Here the success probability p indicates the probability of the mixture distribution being 0. Thus we get in general a distribution of the form

$$Y \begin{cases} = 0 & \text{, with probability } p \\ \sim F & \text{, with probability } 1-p, \end{cases}$$

where F is a continuous distribution. Let $X \sim F$, then one can see that

$$\mathbb{E}\left[Y\right] = (1-p)\mathbb{E}\left[X\right]$$

holds. With small effort one can further derive

$$\mathbb{V}\text{ar}\left[Y\right] = (1-p)\mathbb{E}\left[X^2\right] - (1-p)^2\mathbb{E}\left[X\right]^2.$$

Zero-Inflated Gaussian Distribution

Definition 3.4: *Let $Y = 0$ with probability ν and $Y \sim \mathcal{N}(\mu, \sigma)$ with probability $1-\nu$. Then Y has a* ***zero-inflated Gaussian distribution*** *, where $\nu \in (0,1)$ is the probability for $Y = 0$. We then write $Y \sim ZIG(\mu, \sigma, \nu)$.*

For $\nu = 0$ we get the Gaussian distribution, for $\nu = 1$ we get a degenerated distribution, localized at 0. It is easy to see that the density of $Y \sim ZIG(\mu, \sigma, \nu)$ is given by

$$f_Y(y|\mu, \sigma, \nu) = \begin{cases} \nu & , y = 0 \\ (1-\nu)\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} & , y \neq 0 \end{cases}.$$

As seen for the general case it is straight forward to show that

$$\mathbb{E}\left[Y\right] = \int_{-\infty}^{+\infty} y \cdot f_Y(y|\mu, \sigma, \nu)\mathrm{d}y = (1-\nu)\mu$$

and

$$\mathbb{V}\text{ar}\,[Y] = \int_{-\infty}^{+\infty} (y - (1-\nu)\mu)^2 \cdot f_Y(y|\mu,\sigma,\nu)\mathrm{d}y = (1-\nu)(\mu^2+\sigma^2) - (1-\nu)^2\mu^2$$

for all valid choices of μ, σ and ν.

Yet the zero-inflated Gaussian distribution has not been implemented in the gamlss package. But in this case the zero-inflated Gaussian distribution can easily be implemented following the instructions in (Stasinopoulos et al., 2008, Section 4.2). The required derivatives of the log-likelihood stay the same for μ and σ since the new parameter ν enters additively into the log-likelihood:

$$\begin{aligned} l(y|\mu,\sigma,\nu) &:= \ln(f_Y(y|\mu,\sigma,\nu)) \\ &= \begin{cases} \ln(\nu) & , y = 0 \\ \ln(1-\nu) - \ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 & , y \neq 0 \end{cases}. \end{aligned}$$

The new required derivatives for ν as well as cross-derivatives can be calculated easily. The further needed probability and quantile functions as well a random number generator for the ZIG-distribution can be derived similarly. In Figure 3.3 we give two examples of densities of zero-inflated Gaussian distributions.

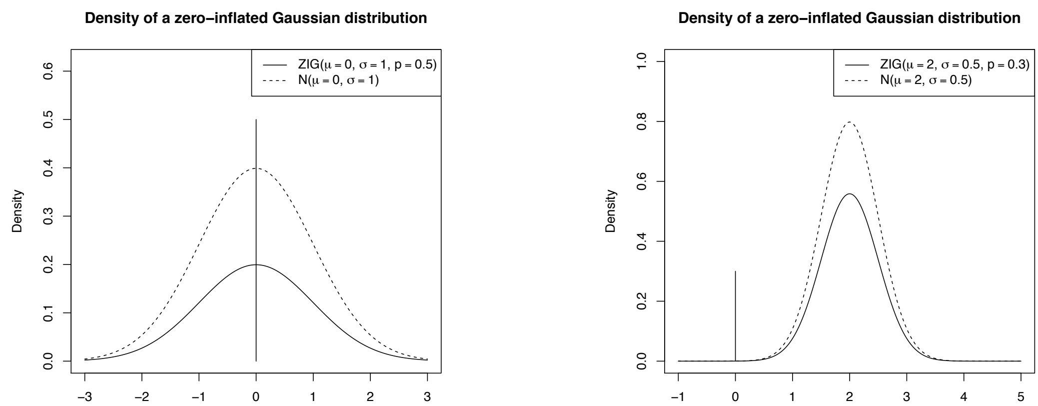


Figure 3.3.: Examples of zero-inflated Gaussian distributions

3.3.1. Zero-Inflated t-Distribution

Definition 3.5: *Let* $Y = 0$ *with probability* τ *and* $Y \sim t(\mu,\sigma,\nu)$ *with probability* $1-\tau$. *Then* Y *has a* ***zero-inflated* $\mathbf{t}$*-distribution*** *, where* $\tau \in (0,1)$ *is the probability for* $Y = 0$. *We then write* $Y \sim ZITF(\mu,\sigma,\nu,\tau)$.

The density of $Y \sim ZITF(\mu, \sigma, \nu, \tau)$ is given by

$$f_Y(y|\mu,\sigma,\nu,\tau) = \begin{cases} \tau & , y = 0 \\ (1-\tau)\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} & , y \neq 0 \end{cases}.$$

One can show that

$$\mathbb{E}\left[Y\right] = \int_{-\infty}^{+\infty} y \cdot f_Y(y|\mu,\sigma,\nu,\tau)\mathrm{d}y = (1-\tau)\mu \qquad \text{for } \nu > 1$$

and

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[Y\right] &= \int_{-\infty}^{+\infty} (y-(1-\tau)\mu)^2 \cdot f_Y(y|\mu,\sigma,\nu,\tau)\mathrm{d}y \\ &= (1-\tau)\left(\sigma^2\frac{\nu}{\nu-2}+\mu^2\right) - (1-\tau)^2\mu^2 \qquad \text{for } \nu > 2.\end{aligned}$$

Like the zero-inflated Gaussian distribution the zero-inflated t-distribution has not been implemented yet in the gamlss package. But similarly to the zero-inflated Gaussian distribution the zero-inflated t-distribution can be implemented using the same technique.

3.4. Skew Exponential Distribution

As readers familiar with reserving might know, there exist lines of business which are very “short-tailed”. That means already after a few development lags no or just very few claims occur. For these lines of business our analyses have shown that even a Gaussian distribution has too heavy tails. The Gaussian distribution would strongly overestimate claims for later development lags and increase reserves. Also observed distributions are sometimes very skewed, putting the use of symmetric distributions into question. Therefore we introduce another distribution. According to the naming in the gamlss package we will call this distribution the skew exponential power type 1 distribution (SEP1 distribution). First we need another distribution to write the SEP1 distribution in a more comfortable way.

Definition 3.6: *Let Y be a random variable with probability density function*

$$f_Y(y|\mu,\sigma,\nu) = \frac{\nu}{2\sigma\Gamma\left(\frac{1}{\nu}\right)} \exp\left\{-\left|\frac{y-\mu}{\sigma}\right|^2\right\} \tag{3.6}$$

for $\mu \in \mathbb{R}$, $\sigma > 0$ and $\nu > 0$. Then Y has a ***power exponential type 2 distribution*** *. We write $Y \sim \mathrm{PE2}(\mu, \sigma, \nu)$.*

There are three special cases:

- For $\nu = 1$ we get a Laplace distribution with mean μ and variance $2\sigma^2$.
- For $\nu = 2$ we get a Gaussian distribution with mean μ and variance $2\sigma^2$.
- For $\nu \to \infty$ we observe a uniform distribution on $[\mu - \sigma, \mu + \sigma]$.

Citing Stasinopoulos et al. (2008) we have

$$\mathbb{E}\left[Y\right] = \mu$$

and

$$\mathbb{V}\mathrm{ar}\left[Y\right] = \sigma^2 \frac{\Gamma\left(\frac{3}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}.$$

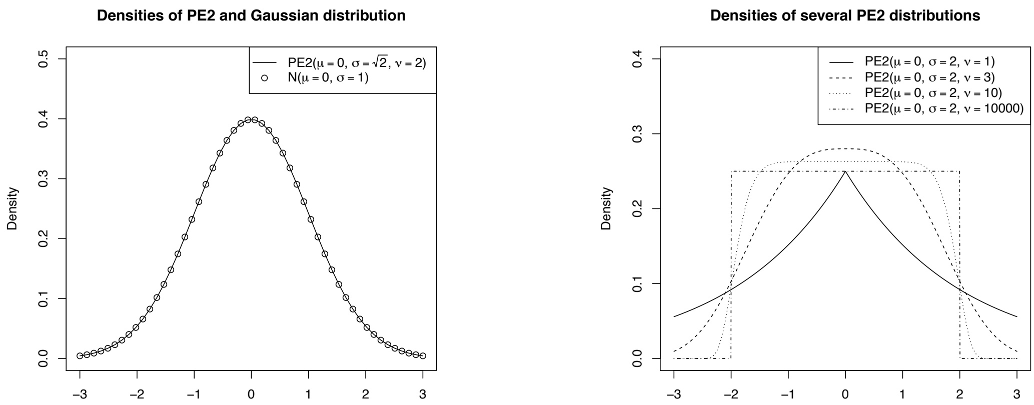


Figure 3.4.: Examples of skew exponential power type 2 distributions

Examples of power exponential type 2 distributions are given in Figure 3.4. For $\nu = 2$ the power type 2 distribution degenerates to a Gaussian distribution with variance $2\sigma^2$. For $\nu \to \infty$ the PE2 distribution converges to a uniform distribution on $[\mu - \sigma, \mu + \sigma]$.

With the PE2 distribution we can now define the SEP1 distribution:

Definition 3.7: *A random variable Y with probability density function*

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{2}{\sigma} f_{Z_2}(z) F_{Z_2}(\nu z) \tag{3.7}$$

*with $y \in \mathbb{R}$, $z = (y - \mu)/\sigma$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\nu \in \mathbb{R}$ and $\tau > 0$ is called **skew exponential power type 1 distribution** , $Y \sim \mathrm{SEP1}(\mu, \sigma, \nu, \tau)$. f_{Z_2} and F_{Z_2} are the density function and cumulative distribution function the power exponential type 2 distribution $Z_2 \sim PE2(0, \tau^{1/\tau}, \tau)$.*

The SEP1 distribution is of Azzalini type 1 (compare Azzalini (1986)). Hence moments can be calculated following Azzalini's approach (see also Stasinopoulos et al. (2008) for details):

$$\begin{aligned}\mathbb{E}[Y] &= \mu + \sigma\mathbb{E}[Z] \\ \mathbb{V}\mathrm{ar}[Y] &= \sigma^2\mathbb{V}\mathrm{ar}[Z] = \sigma^2\left(\mathbb{E}\left[Z^2\right] - \mathbb{E}[Z]^2\right),\end{aligned}$$

where $Z = (Y - \mu)/\sigma)$ and

$$\mathbb{E}[Z] = \mathrm{sign}(\nu)\tau^{1/\tau}\frac{\Gamma\left(\frac{2}{\tau}\right)}{\Gamma\left(\frac{1}{\gamma}\right)}\mathrm{pBEo}\left(\frac{\nu^\tau}{1+\nu^\tau}, \frac{1}{\tau}, \frac{2}{\tau}\right)$$

and

$$\mathbb{E}\left[Z^2\right] = \tau^{2/\tau}\frac{\Gamma\left(\frac{3}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)},$$

with pBEo being the cumulative distribution function Beta distribution (in the gamlss package named *original Beta distribution*). It is important to note that neither the mean nor the variance are simple functions of only one parameter but depend on all four parameters. Thus interpretation of a single parameter is not very easy.
Figure 3.5 gives an impression of what the influences of the different parameters are. μ clearly affects only the location. While σ mainly influences the variance it also has an influence on the mean. ν controls the slope of the density until its maximum. A big ν results in a steep slope with maximum μ. For $\nu \to \infty$, $Y \sim SEP1(0, \sigma, \nu, 1)$ behaves like a $\mathrm{Exp}(1/\sigma)$ distribution (not shown here). While an increasing ν leads to a right-skewed distribution an increasing τ leads to a left-skewed distribution. For $\tau \to \infty$ we see a distribution with the form of a triangular distribution on $[\mu, \mu + \sigma/2)$ and a uniform distribution on $[\mu + \sigma/2, \mu + \sigma)$. Although there may be connections to several other distributions we will present only on one distribution which is a special case of the skew exponential power type 1.

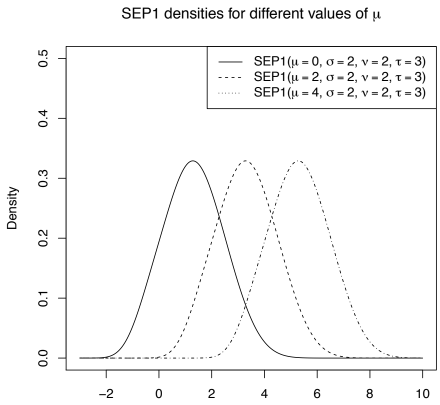
3.4.1. Skew Gaussian Distribution

A special case of (3.7) is $\tau = 2$.

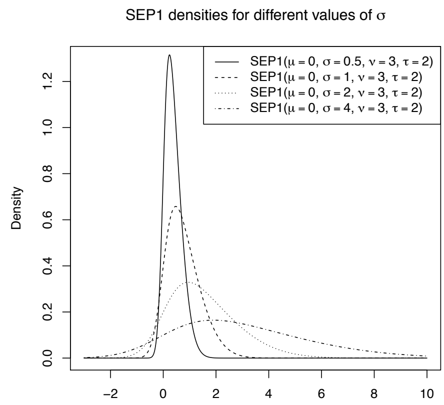
Definition 3.8: *Let $Y \sim \mathrm{SEP1}(\mu, \sigma, \nu, \tau)$ and fix $\tau = 2$, then Y has a **skew normal distribution**.*

The skew exponential power type 1 distribution is defined via the power exponential type 2 distribution. Hence, for $\tau = 2$ we get for the probability density of $Z \sim \mathrm{PE2}(0, 2^{1/2}, 2)$:

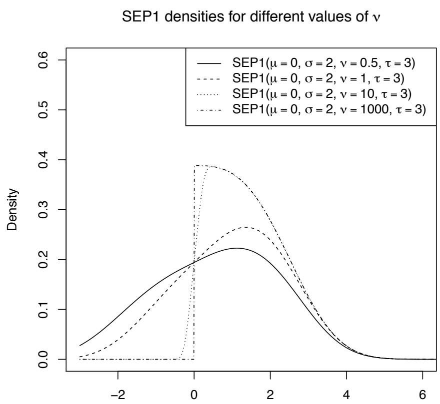
$$f_Z(z|0, 2^{1/2}, 2) = \frac{2}{2 \cdot 2^{1/2}\Gamma\left(\frac{1}{2}\right)}\exp\left\{-\left|\frac{z-0}{2^{1/2}}\right|^2\right\} = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\} = \phi(z),$$



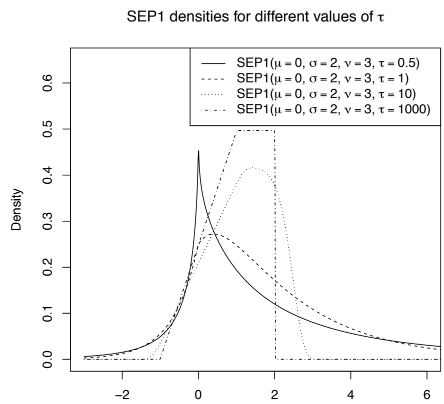
(a) μ only affects the location (and thus the mean)



(b) σ not only affects the variance but the location as well



(c) The bigger ν is the steeper the curve until its maximum



(d) τ controls the skewness of the distribution

Figure 3.5.: Influence of the different parameters on the SEP1 distribution

where ϕ is the density of a standard Gaussian distribution. Then, following O'Hagan and Leonard (1976), $Y \sim \text{SEP1}(0, 1, \nu, 2)$ has a standard skew normal distribution $Y \sim \text{SN}(0, 1, \nu)$ with density

$$f_Y(y|0, 1, \nu) = 2\phi(y)\Phi(\nu y),$$

where Φ is the distribution function of a standard Gaussian distribution. Using the transformation $y \mapsto \frac{y-\mu}{\sigma}$ it follows that $Y \sim \text{SN}(\mu, \sigma, \nu)$ has a skew normal distribution. The skew normal distribution still belongs to the class of Azzalini's type 1 distributions, such that the mean and variance can be calculated as

$$\mathbb{E}\left[Y\right] = \mu + \sigma \cdot \text{sign}(\nu)\sqrt{\frac{2\nu^2}{\pi(1+\nu^2)}}$$

and

$$\mathbb{V}\text{ar}\left[Y\right] = \sigma^2 \left(1 - \frac{2\nu^2}{\pi(1+\nu^2)}\right),$$

compare also Stasinopoulos et al. (2008).

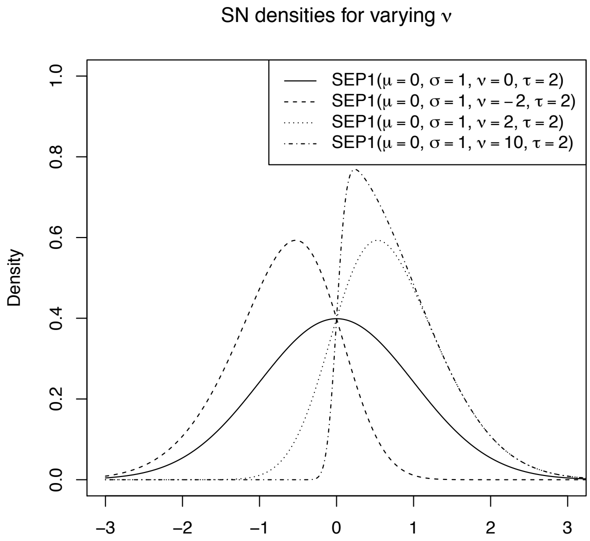


Figure 3.6.: Density plot of skew normal distributions

Figure 3.6 shows the density of skew normal distributions with different ν values. For $\nu = 0$ the skew normal distribution is an ordinary Gaussian distribution. Extreme values for ν in both directions lead to more skewed distributions.

3.5. Summary

The Gaussian and t-distribution are standard and well studied distributions. Especially regarding computing time their implementation is very fast. The zero-inflated Gaussian and zero-inflated t-distribution had to be implemented manually. But since we could make use of the already implemented continuous version calculation time is very acceptable as well. The skew exponential power type 1 distribution is part of the gamlss package and hence did not need to be implemented. However, no closed form for the distribution function is known to us or the authors of the gamlss package. Hence numerical integration is used, making the calculation (very) time-consuming.

4. Application of GAMLSS

After having introduced some theory about GAMLSS this and the following sections are dedicated to model fitting and simulation studies.
We apply GAMLSS to an insurance portfolio of 6 lines of business. For five lines of business (LoB 1 - 5) with continuous distribution we will examine dependencies in a later chapter. For those five lines of business data is available for 21 years on a yearly basis, such that there are 231 observations for each line of business. For one line of business (LoB 6) a zero-inflated distribution is used. For this line of business which is analyzed in this section, only 20 years on a yearly basis of data is available, such that there are 210 observations.

While the chain ladder method needs cumulative payments, the GAMLSS is indifferent to which data is used. We could use cumulative or incremental payments as well, but there are several problems when using them. For instance data is not inflation adjusted, possibly leading to a natural increase of paid losses among calendar years. Another point is that business changes over time. The insurance company could write more policies and in general this would lead to higher payments (in total, not necessarily per contract). Or the other way around, leading to decreasing payments. To detrend data we use a ratio which we call paid-to-premium ratio.

Definition 4.1: *Let*

$$\mathcal{S} = \{S_{i,k} | i = 1, \ldots, n,\ k = 1, \ldots, n + 1 - i\}$$

be a triangle with incremental payments. Let

$$\mathcal{P} := \{P_{i,k} | i = 1, \ldots, n,\ k = 1, \ldots, n + 1 - i\}$$

be a triangle with cumulative earned premiums. Then

$$PtP_{i,k} := \frac{S_{i,k}}{P_{i,k}} \quad i = 1, \ldots, n,\ k = 1, \ldots, n + 1 - i \tag{4.1}$$

is called (incremental) ***paid-to-premium ratio****.*

Because premiums are earned only once at the beginning of the contract, a triangle with incremental premiums would only have premiums in the first column and zeroes in all other columns. A triangle with cumulative earned premiums then means to carry forward the premium of the first column, compare Table 4.1.

Accident Year i	Development Lag k 1	2	3	4
1	1000	0	0	0
2	1100	0	0	
3	1800	0		
4	1200			

(a) Incremental earned premiums

Accident Year i	Development Lag k 1	2	3	4
1	1000	1000	1000	1000
2	1100	1100	1100	
3	1800	1800		
4	1200			

(b) cumulative earned premiums

Table 4.1.: Triangles with incremental and cumulative earned premiums

Since premiums stay constant in each row, we can drop the index k, so that P_i is the earned premium for accident year i for all development lags.

Note: In this thesis we will only use earned premiums and no written premiums. We use the term 'premiums' henceforth and mean earned premiums. Readers familiar with insurance products notice that premiums can be earned also at later development lags for certain lines of business. However, data used in this thesis includes only lines of business for which the whole premium is earned at the beginning of the contract and premium triangles are of the form shown in Table 4.1. $\diamondsuit$

By using the premium as an exposure, certain unusual observations for a triangle with incremental payments may vanish. E.g. for the line of business with premiums as in Table 4.1, if there were high payments for accident year 3 at both development lags, this would no longer be suspicious, since a much higher premium has been earned for this accident year. Higher premiums in general refer to more written policies and hence more claims can occur.
Another in insurance companies more often regarded ratio, is the loss ratio. Besides the raw payment for the claim, the loss ratio includes possible adjustment expenses. But due to lack of availability of adjustment expenses in our data set, we will not consider the loss ratio.

We then need one or more distributions for the PtP ratios and a set of covariates. The set of covariates is rather small and consists of development lag, accident year and implied be them calendar year.
Which distribution should be used is a bit more complicated. Neglecting for a moment the distributions introduced in Chapter 3, the gamlss package which we use for model fitting provides more than 30 distributions. Clearly, a preselection has to be done since fitting models for all distributions is neither possible nor reasonable due to restrictions of

the distributions (i.e. discrete/continuous distributions or different support). Hence we need to find a set of suitable distributions at the beginning.
The first question we have to answer is whether to use a discrete, continuous or mixed distribution. Recall that for the line of business analyzed in the following there are 210 observations from 20 years available. An excerpt of the data set is shown below.

```
> lob6.data[1:5,]

   DevLag AYear CYear PtP.Ratio
7       1  1992  1992 0.3224859
8       1  1993  1993 0.3539408
9       1  1994  1994 0.3639705
10      1  1995  1995 0.3452195
11      1  1996  1996 0.3247802
```

While a discrete distribution is not reasonable for PtP ratios we count the observed zeroes to choose between a continuous and a mixed distribution.

```
> sum(lob6.data$PtP.Ratio == 0)

[1] 24
```

24 out of 210 observations are zero (or in relative terms roughly 11.4%). That clearly is too much to assume a continuous distribution for Y. We need a zero-inflated distribution and some of them have been implemented in the gamlss package like a zero-adjusted inverse Gaussian distribution. But observe

```
> sum(lob6.data$PtP.Ratio < 0)

[1] 46
```

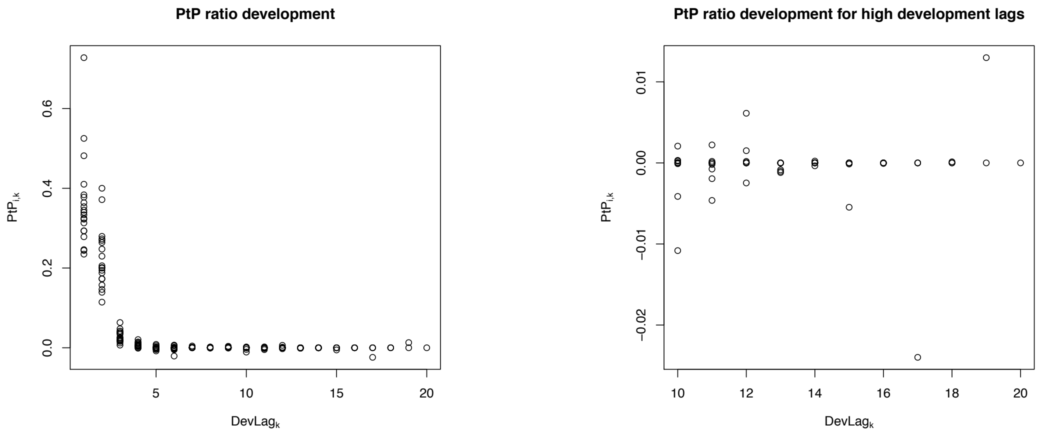
i.e. 46 of 210 observations are negative. This disqualifies mixed distribution with support on the positive real numbers only. Thus we need a mixed distribution with enough mass on 0 and support on the whole real line. The zero-inflated Gaussian and zero-inflated t-distribution in Chapter 3 are two examples for that kind of distribution. Hence we can use both of them but need to decide which of them is the preferred one later.

4.1. Modeling the Location Parameter

Having found a set of distributions we examine influences of the different covariates.
We start with influences on the location parameter. This is particularly of interest since for both the zero-inflated Gaussian and the zero-inflated t-distribution the location parameter μ is closely related to the mean. The mean can be used as an estimator for unknown future ratios and hence payments.

Development Lag

We expect a strong dependence of PtP ratios on development lags because that is what we know from other methods like chain ladder. And we are mainly interested in how claims develop in the future.



(a) PtP ratios decrease for increasing development lags

(b) Detailed excerpt of (a) for high development lags

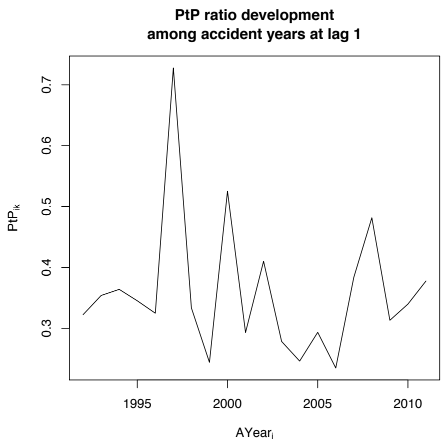
Figure 4.1.: Development of PtP ratios for LoB 6. Although PtP ratios seem to be almost constant for higher lags in (a) they are not as one can see in (b)

Figure 4.1 shows PtP ratios plotted against development lag. We see a strong decrease in PtP ratios and even for higher development lags (Figure 4.1(b)) ratios are not constant. Obviously development lags have a high influence and we will incorporate this by using a polynomial on development lag. The degree has to be specified during residual analyses, but as starting point we set the degree to 5.

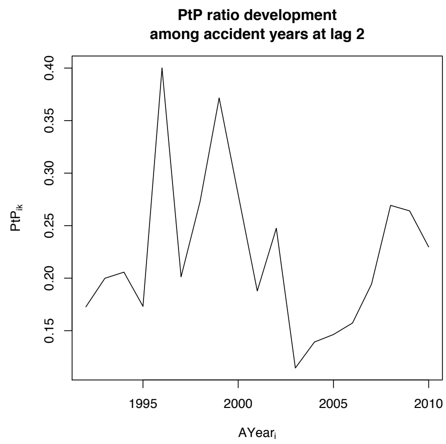
Note: Other models (compare e.g. Shi and Frees (2011)) let the development lag enter as a factor. We don't follow this approach since this would increase the number of parameters significantly. While we start with 6 parameters (polynomial plus intercept), a model with development lags as factors would have 20 parameters. ♢

Accident Year

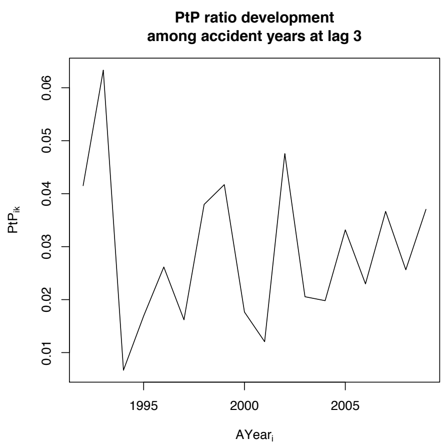
Although hoping to have excluded most of accident year effects by working on PtP ratios this has to be verified.
Figure 4.2 shows no obvious trend among accident years, only the scale becomes smaller for increasing development lags. There are some extreme values at each lag but not for the same accident year. We therefore do not use accident year at this point of the process. We will investigate in this again after we fit a first model and check residuals.



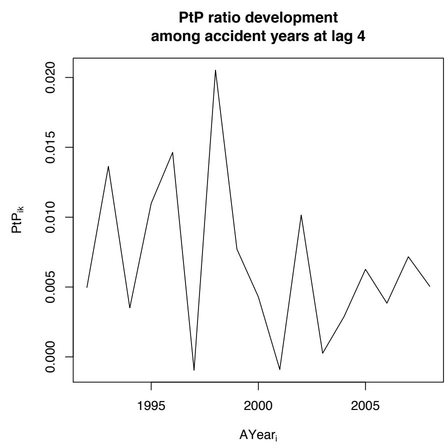
(a) PtP ratios among accident year at development lag 1



(b) PtP ratios among accident year at development lag 2



(c) PtP ratios among accident year at development lag 3



(d) PtP ratios among accident year at development lag 4

Figure 4.2.: Development of PtP ratios among accident year for the first four development lags

Calendar Year

Since calendar year effects are even harder to detect than accident year effects we do not include the calendar year at this stage. Like for accident year we will check for an influence later on the residuals of the model.

4.2. Model Fitting

The gamlss package allows modeling of up to four parameters in its current implementation. We could continue to examine influences of the covariates on the scale and shape parameters in a similar way. However, we do not favor this. Then it would very difficult to relate certain observations in the model checking process to a specific parameter. Another reason is that when modeling the first or the first two parameters well, there may be no necessity to model more parameters at all. For this we have to check residuals and thus need to fit a model first. And finally we want to find a good model for the mean which depends strongly on the location parameter μ for most distributions. To not entirely exclude other parameters, a constant will be estimated for them.

The studies for development lag and accident year influence suggest a model of the form

$$Y \sim \text{ZIG}(\mu = poly(DevLag_k, 5), \sigma = 1, \nu = 1) \tag{ZIG 1}$$

or

$$Y \sim \text{ZITF}(\mu = poly(DevLag_k, 5), \sigma = 1, \nu = 1, \tau = 1). \tag{ZITF 1}$$

Note that while for the ZIG ν models the probability of Y being zero, τ models that probability for the ZITF.
First we fit a model with zero-inflated Gaussian distribution.

```
> ptp.gamlss.zig <- gamlss(PtP.Ratio ~ poly(DevLag, 5), sigma.formula=~1,
+ nu.formula=~1, family=ZIG(), method=mixed(), data=lob6.data)

GAMLSS-RS iteration 1: Global Deviance = -454.9138
GAMLSS-CG iteration 1: Global Deviance = -454.9138
```

Some notes on the options used here:

- `PtP.Ratio` stands for the paid-to-premium ratios, `DevLag` for the development lag
- The `poly()` function chooses orthogonal polynomials by default
- The formulas for `sigma.formula` and `nu.formula` don't need to be stated if they are not different from a constant
- `method` allows choosing either a Cole-Green algorithm, a Rigby-Stasinopolous algorithm or a mixture of both of them, compare (Rigby and Stasinopoulos, 2005, Section 5). While results using either the CG or RS algorithm don't differ much if they converge, sometimes only one of them can be used to assure convergence at all.

```
> summary(ptp.gamlss.zig)

*******************************************************************
Family:  c("ZIG", "ZeroInflGaussian")

Call:
gamlss(formula = PtP.Ratio ~ poly(DevLag, 5), sigma.formula = ~1,
    nu.formula = ~1, family = ZIG(), data = lob6.data, method = mixed())
```

```
Fitting method: mixed()

-------------------------------------------------------------------
Mu link function:  identity
Mu Coefficients:
                  Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)        0.05676    0.003628   15.648  8.766e-37
poly(DevLag, 5)1  -1.02813    0.059102  -17.396  3.648e-42
poly(DevLag, 5)2   0.94717    0.056125   16.876  1.424e-40
poly(DevLag, 5)3  -0.75042    0.063017  -11.908  3.620e-25
poly(DevLag, 5)4   0.47802    0.071075    6.726  1.723e-10
poly(DevLag, 5)5  -0.22550    0.066491   -3.392  8.347e-04

-------------------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
             Estimate  Std. Error  t value    Pr(>|t|)
(Intercept)    -3.043     0.05185   -58.69  7.987e-132

-------------------------------------------------------------------
Nu link function:  identity
Nu Coefficients:
             Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)    0.1143     0.02195    5.205  4.615e-07

-------------------------------------------------------------------
No. of observations in the fit:  210
Degrees of Freedom for the fit:  8
      Residual Deg. of Freedom:  202
                      at cycle:  1

Global Deviance:     -454.9138
            AIC:     -438.9138
            SBC:     -412.1369
*******************************************************************
```

We assume the reader is familiar with model fitting in R for linear models or generalized linear models and hence is familiar with the `summary` function for these models. Having a look at the summary for the GAMLSS we see that all parameters are significant. Beside parameter estimates the corresponding link functions for the different parameters are printed. This is important when comparing parameter estimates of different models with possibly different link functions. For this model 8 parameters have been used: 1 for the intercept + 5 for polynomial + 1 for σ + 1 for ν.
We have a look at residuals to check model assumptions. In the gamlss package normalized

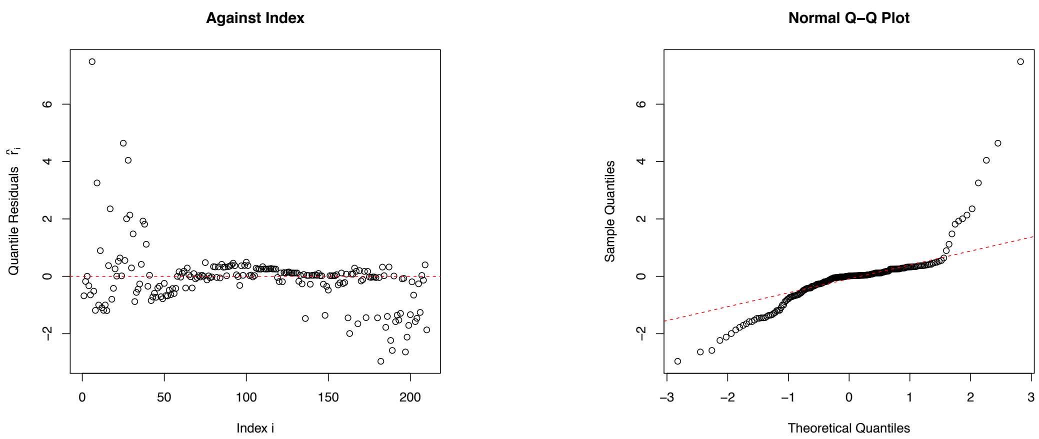
quantile residuals are used. For observation $i = 1, \dots, n$ **normalized quantile residuals** are defined as

$$\hat{r}_i := \Phi^{-1}(u_i),$$

where

$$u_i := F(y_i|\hat{\boldsymbol{\theta}}^i).$$

In words, observation y_i is transformed to $u_i \in [0, 1]$ using the cumulative distribution function F of the underlying distribution with estimated parameters from the GAMLSS. Then the inverse cumulative distribution function of a standard normal variate is applied to u_i to get $\hat{r}_i$. While this seems to be not very intuitive at first sight, it has the advantage that $\hat{r}_i$, $i = 1, \dots, n$, should then be approximately standard normal distributed. Any deviation from this can be seen in a Q-Q normal plot.



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure 4.3.: Both plots show a very bad fit for the ZIG model

Figure 4.3 shows a plot of normalized quantile residuals against the index and a Q-Q normal plot of them. Obviously this shows a very bad fit. The normalized quantile residuals are neither normal distributed nor does the plot against the index look well, i.e. shows no pattern.

Note: To fit a GAMLSS to triangle data, data has to be aligned differently. We chose to sort the data by development lag first and accident year afterward. An example for small 3x3 triangle is shown in Table 4.2.

Index	DevLag	AYear	PtP.Ratio
1	1	1	0.52
2	1	2	0.55
3	1	3	0.59
4	2	1	0.31
5	2	2	0.23
6	3	1	0.12

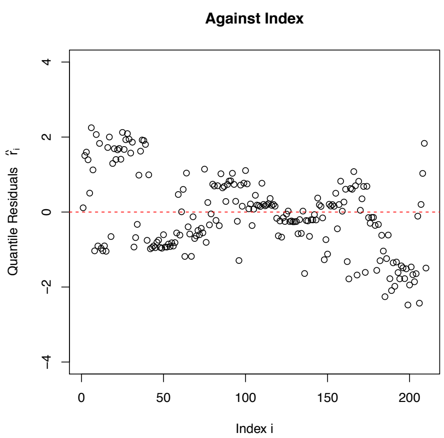
Table 4.2.: Example how the index i has to be read (3x3 triangle)

Hence the **index** i runs among the columns of the triangle for one development lag after another. Looking at Figure 4.3(a), the first 21 observations refer to development lag 1, the next 20 to development lag 2, etc. ♢

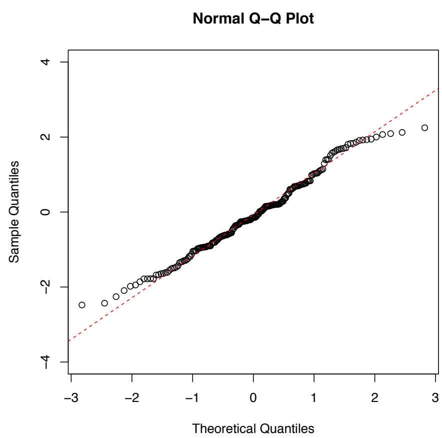
We fit the same model with zero-inflated t-distribution and compare the residual plots in Figure 4.4.

```
> ptp.gamlss.zitf <- gamlss(PtP.Ratio ~ poly(DevLag, 5), sigma.formula=~1,
+ tau.formula=~1, family=ZITF(), method=RS(100),
+ control=gamlss.control(trace=F), data=lob6.data)
```

A note on the `gamlss` options: We used the RS algorithm (with up 100 iterations) only and deactivated printing the deviance at each iteration by setting `control = gamlss.control(trace = F)`.



(a) Plot of normalized quantile residuals against index



(b) Q-Q normal plot of normalized quantile residuals

Figure 4.4.: Residual plots of the ZITF 1 model show better a fit than for the ZIG 1 model

Figure 4.4 shows the improvement by using a zero-inflated t-distribution. The Q-Q normal plot for the ZIG indicated that PtP ratios follow a distribution with heavier tails than the Gaussian distribution. The zero-inflated t-distribution captures this behavior and the Q-Q normal plots looks better. But the plot of normalized quantile residuals against the index still does not look well as patterns are observable.

Remark: The algorithms respect rules for estimated parameters up to a certain level. E.g. the degree of freedom of a t-distribution should always be positive. This restriction is implemented and no negative degree will be estimated. But we also need to ensure existence of the first two moments and hence a degree of freedom of at least 2. Checking the parameter estimates of the ZITF model we see that this is not fulfilled. The degrees of freedom in this model are

```
> ptp.gamlss.zitf$nu.fv[[1]]
```

```
[1] 0.6569353
```

which is even smaller than 1 and not even the mean exists. Thus when observing such situations we have to fix ν manually at $2+\varepsilon$, $\varepsilon > 0$, such mean and variance exist. We decided to fix ν in these cases at $\nu = 2.1 =: 2^+$. The model becomes

$$Y \sim \text{ZITF}(\mu = poly(DevLag_k, 5), \log(\sigma) = 1, \nu := 2^+, \tau = 1). \qquad \text{(ZITF 2.1)}$$

To fix the degrees of freedom in R, we use the model

```
> ptp.gamlss.zitf2.1 <- gamlss(PtP.Ratio ~ poly(DevLag, 5), sigma.formula=~1,
+ tau.formula = ~1, family=ZITF(), method=RS(100),
+ control=gamlss.control(trace=F), nu.start=2.1, nu.fix=T, data=lob6.data)
```

which leads to different residual plots (see Figure 4.5). Especially the Q-Q normal plot shows the bad influence of fixing the degree of freedom at a higher level than estimated. Though note that it still looks better than using a zero-inflated Gaussian distribution.

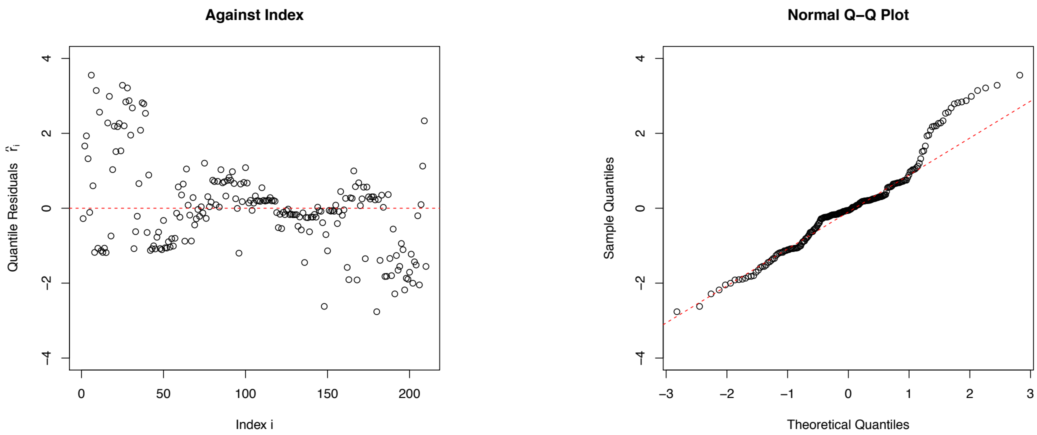
$\diamondsuit$

Figure 4.5(a) shows big differences for the variation among the index. While the spread for the first 100 indices is big, it is very small for indices 100 - 140 and gets bigger afterward. Thus we need to model the second parameter σ which is able to capture this. We fit a third model

$$\begin{aligned} Y \sim \text{ZITF}\big(&\mu = poly(DevLag_k, 5), \log(\sigma) = poly(DevLag_k, 5), \\ &\nu := 2^+, \tau = 1\big), \end{aligned} \qquad \text{(ZITF 2.2)}$$

where a polynomial of degree 5 is used for σ with a log link function. The log link ensures that we only get positive fitted values for σ. Furthermore we fix ν at $2^+ := 2.1$.

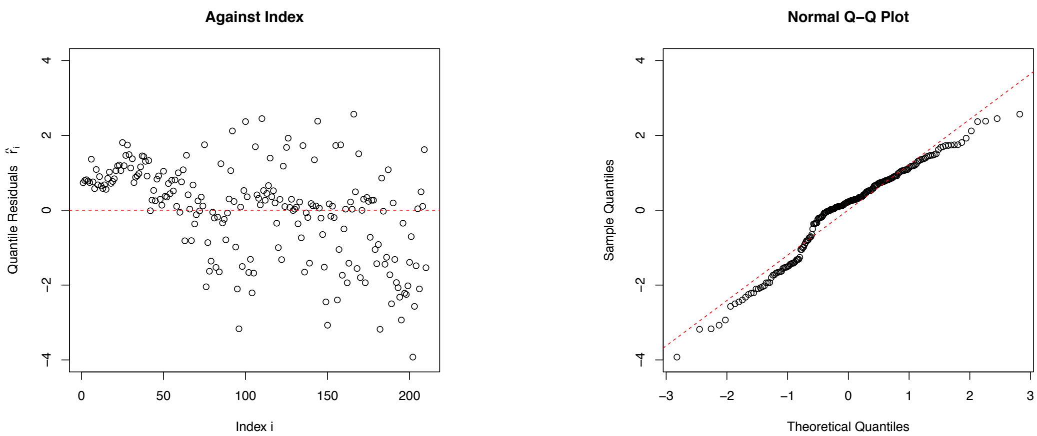
```
> ptp.gamlss.zitf2.2 <- gamlss(PtP.Ratio ~ poly(DevLag, 5),
+ sigma.formula=~poly(DevLag, 5), tau.formula=~1, family=ZITF(), method=RS(100),
+ control=gamlss.control(trace=F), nu.start=2.1, nu.fix=T, data=lob6.data)
```



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure 4.5.: Residual plots of the ZITF 2.1 model with fixed degree of freedom at $\nu = 2.1$



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure 4.6.: Residual plots of the ZITF 2.2 model

The residual plots are shown in Figure 4.6 and see the improvement by modeling the scale parameter. Residuals with index greater than 70 seem to scatter around zero uniformly. Also the Q-Q normal plot shows a better fit, especially in the upper tail.

Note: While in this case the improvement can be visualized well this might not always be the case. We then need a statistics to measure goodness of fit. For GAMLSS goodness of fit is measured in terms of the **global deviance**,

$$GD = -2l(\hat{\boldsymbol{\theta}}) = -2\sum_{i=1}^{n} l(\hat{\boldsymbol{\theta}}^i) = -2\sum_{i=1}^{n} \ln(f(y_i|\hat{\boldsymbol{\theta}}^i)).$$

Note that here the complete log-likelihood including all terms is used while for GLM it is not in general. Then goodness of fit is then measured by the **Schwarz Bayesian criterion**,

$$SBC = GD + \ln(n) \cdot df,$$

where df is the total effective degree of freedom used in the model. The model with lower SBC is then the better one. In contrast to the AIC (Akaike's information criterion) the SBC penalizes the number of parameters heavier through the second term. Since the number of parameter can grow rapidly for GAMLSS, we think this is more suitable for our purposes. $\diamondsuit$

For the two models the SBC is

```
> ptp.gamlss.zitf2.1$sbc
```

```
[1] -674.4351
```

and

```
> ptp.gamlss.zitf2.2$sbc
```

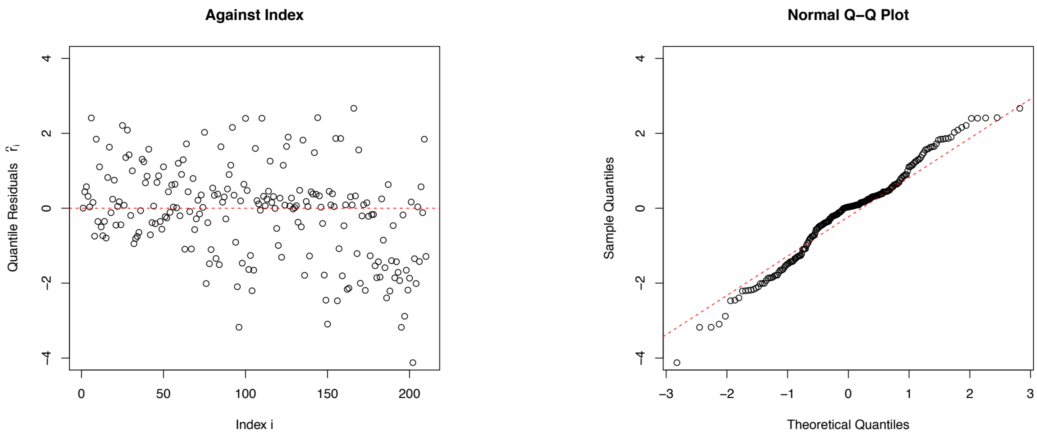
```
[1] -1032.481
```

confirming that the model with explicitly modeled scale parameter is much better. But the model ZITF 2.2 still shows a lack of fit, especially for the first indices, i.e. the first development lags. We therefore add factors for the first four development lags to the model. Analyses showed that in this case parameters for the polynomial on development lag are not significant so we drop them. The model for the PtP ratios is then

$$\begin{aligned} Y \sim \text{ZITF}(\mu = &\; DevLag_k + (DevLag_k = 1) + (DevLag_k = 2) \\ &+ (DevLag_k = 3) + (DevLag_k = 4), \\ \log(\sigma) = &\; poly(DevLag_k, 5), \\ \nu = &\; 2^+, \\ \tau = &\; 1), \end{aligned} \tag{ZITF 2.3}$$

where $(DevLag_k = x)$ is a factor on development lag for lag x:

$$(DevLag_k = x) = \begin{cases} 1 & , \text{if } k = x \\ 0 & , \text{else} \end{cases}.$$



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure 4.7.: Residuals plots of the ZITF 2.3 model

One may surely argue that using factors for all development lags could lead to an even better fit. Then to go further one could use factors for all development lags for the scale parameter as well and even add factors for the accident year. But then we were in a situation we wanted to omit because we would have a model with more than 50 parameters on 210 observations.
The plots in Figure 4.7 show a better fit than for model ZITF 2.2 for the first development lags. The latest model ZITF 2.3 has

```
> ptp.gamlss.zitf2.3$df.fit

[1] 13
```

degrees of freedom which is the same as for model ZITF 2.2

```
> ptp.gamlss.zitf2.2$df.fit

[1] 13
```

But obviously there is still some need to improve the model when looking at the variation, which is bigger for later accident years.

So far we haven't analyzed the pattern of zeroes in the triangle at all. Recall that for the ZITF τ models the *probability* of Y being zero. Estimating a constant means to assign the same probability to all cells in the triangle. This does not make much sense, since naturally there won't be a lot of zeroes for the first development lags but possibly more for higher development lags.
However, simple counting of zero-observations among development lags is not useful to

get an idea of the behavior of τ. Instead we divide the number of zero-observations by the total number of observations per development lag. In formulas this is

$$z_k = \frac{\sum_{j=1}^{n+1-k} \mathbb{1}_{PtP_{j,k}=0}}{n+1-k}, \quad k = 1, \ldots, n.$$

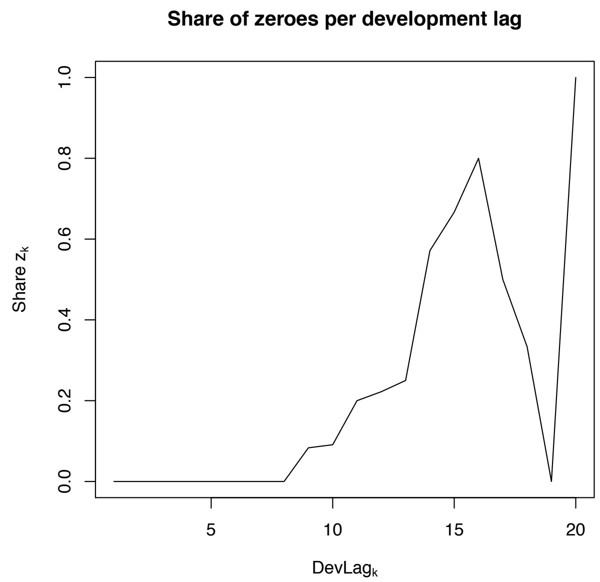


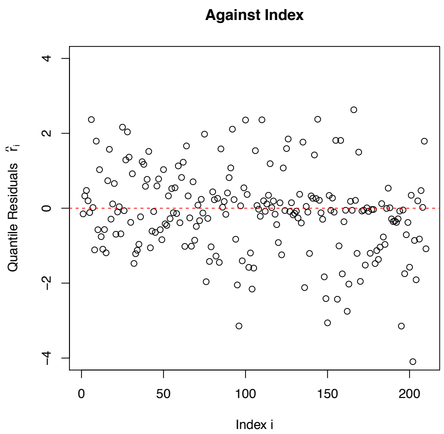
Figure 4.8.: Zeroes appear only for later accident years

Figure 4.8 shows a plot of z_k against development lag. No zeroes occur for lags less than 9 and afterward they seem to follow a quadratic trend (neglecting the latest development year which has only one observation). We incorporate this by adding a new covariate called 'adjusted development lag', which is defined as

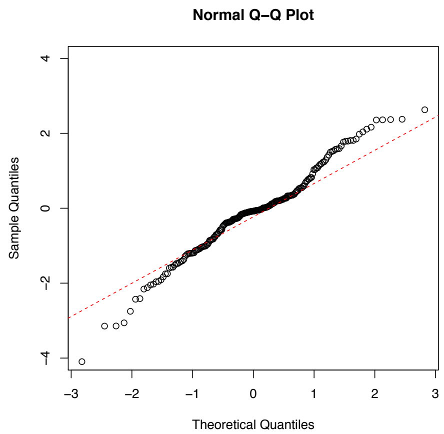
$$Adj.DevLag_k = \begin{cases} 0 & , DevLag_k \leq 8 \\ DevLag_k - 8 & , DevLag_k > 8 \end{cases} \tag{4.2}$$

We then use a polynomial of degree 2 on this covariate to model τ:

$$\begin{aligned} Y \sim \text{ZITF}(\mu = {} & DevLag_k + (DevLag_k = 1) + (DevLag_k = 2) \\ & + (DevLag_k = 3) + (DevLag_k = 4), \\ \log(\sigma) = {} & poly(DevLag_k, 5), \\ \nu = {} & 2^+, \\ logit(\tau) = {} & poly(Adj.DevLag_k, 2)\big), \end{aligned} \tag{ZITF 2.4}$$



(a) Plot of normalized quantile residuals against index



(b) Q-Q normal plot of normalized quantile residuals

Figure 4.9.: Residuals plots of the ZITF 2.4 model

where *logit* states that a logit-link is used for τ. This is necessary to ensure that all fitted values for τ are between 0 and 1.
Figure 4.9 shows the residuals plots for model ZITF 2.4. It is not easy to see the improvement in the plot of residuals against the index. Here it is useful to have the SBC, which dropped by more than 50 and shows an improvement.

```
> ptp.gamlss.zitf2.4$sbc - ptp.gamlss.zitf2.3$sbc

[1] -54.30414
```

On the other side the Q-Q normal plot looks worse than before. To improve the model further, recall that we have data which is normally analyzed in *triangles*. We can use this information an arrange residuals in a triangle which can help to identify patterns which are not observable in the plot against the index or the Q-Q normal plot. Such a plot is shown in Figure 4.10.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1992	−0.1506	−0.6939	0.7715	−0.1414	1.9782	1.0772	1.5369	−0.9163	0.3724	0.2122	−2.4289	2.6285	−0.0845	0.1233	−0.3558	−0.0465	0.3462	0.1975	0.0226	−1.0824
1993	0.3295	−0.0986	1.5167	1.1304	−1.9614	2.1109	0.0724	0.1456	−0.3922	−0.1224	−1.0057	0.2099	−1.2024	−0.7677	−0.3423	−1.748	−4.0957	−0.8228	1.7901	
1994	0.4743	0.0407	−1.0557	−0.3866	−0.2719	0.2311	−0.0262	−1.243	1.7656	−0.2949	1.8103	−1.2055	−0.0331	−0.9661	−0.3784	−0.7078	−0.8599	0.4721		
1995	0.1986	−0.6836	−0.6108	0.8215	−1.4206	−0.8274	−0.2146	1.0754	−2.1185	−1.8321	−1.7519	1.4975	−0.0331	−0.0058	−0.2829	−0.3797	−1.9095			
1996	−0.1156	2.1642	−0.089	1.2288	−1.0299	−2.048	2.3602	−0.0608	0.0483	−2.4124	−0.3596	−1.9533	−1.4764	0.5359	−0.0751	−1.5744				
1997	2.3674	−0.0649	−0.6464	−1.0187	0.4375	−3.1433	0.1988	1.5932	−0.1029	−3.0593	−0.0534	−0.074	−1.1295	0.0119	−3.1459					
1998	0.0192	1.29	0.5952	1.6634	0.2273	0.0651	−0.0822	1.848	−1.2074	0.3418	−2.7515	−0.0503	−1.3662	−0.2857						
1999	−1.1113	2.0381	0.7809	0.3294	−1.2766	−1.4044	0.1145	−0.0873	0.3278	−0.055	−2.0246	−1.5177	−1.0381							
2000	1.7917	1.3639	−0.5722	−0.2565	0.2636	0.5452	0.3467	0.1479	0.2656	0.272	0.1845	0.0098								
2001	−0.5733	−0.377	−0.8399	−1.0137	−1.4455	2.3554	1.1897	−0.177	1.423	−0.1038	−0.0554									
2002	1.0304	0.9207	1.032	0.707	1.5844	0.3675	0.0166	−0.1365	0.2585	1.8109										
2003	−0.7582	−1.4738	−0.4163	−0.8536	0.0278	−1.5783	0.1928	−0.0776	2.3753											
2004	−1.093	−1.2143	−0.4575	−0.4854	0.1796	−1.1937	−0.159	−0.2594												
2005	−0.5694	−1.1214	0.3298	0.0832	−0.1626	−2.1576	−0.4372													
2006	−1.19	−0.9618	−0.2781	−0.3314	0.409	−1.597														
2007	0.7349	−0.2286	0.5252	0.2385	0.8177															
2008	1.5733	1.2418	−0.1202	−0.1276																
2009	−0.2906	1.1721	0.5463																	
2010	0.1152	0.5854																		
2011	0.6601																			

1%− quantile
99%− quantile

Figure 4.10.: Residuals arranged in triangular form help to identify outliers

In Figure 4.10 residuals are arranged in triangular form. Recall that normalized quantile residuals are shown here. The coloring indicates whether the residuals fall below the (theoretical) 1%-quantile (light grey), above the 99%-quantile (dark grey) or in between (white). We can't see any obvious patterns, i.e. increasing residuals among accident years or calendar years. But for accident year 1993 at development lag 17 we see the huge negative outlier. Having a look at the underlying data set we detect a very unusual PtP ratio. Thus we decide to model this ratio by a factor on this cell. Note that we could have excluded the data entry as well. The second observation we make is that accident year 1997 has 3 negative outliers. Since residuals for this accident year take positive and negative values, a factor on accident year 1997 for σ is used.
Furthermore Figure **??** shows unusually low variances for development lags 3 and 4 a big variance for lag 6. We include this by factors on those lags for σ. Note that we focus on the first development lags more than on later ones. This is because they influence the prediction of future claims more than later development lags and thus it is important to model them well.

We added all new covariates successively and checked residuals and significance of parameters at each step. Hence some covariates dropped out like while some other entered the model. Our final model for this line of business is

$$\begin{aligned} Y \sim \mathrm{ZITF}\Big(\mu &= poly(DevLag_k, 2) + (DevLag_k = 1) + (DevLag_k = 2) \\ &\quad + (DevLag_k = 3) + (DevLag_k = 4) \\ &\quad + (DevLag_k = 17 \,\&\, AYear_i = 1993), \\ \log(\sigma) &= poly(DevLag_k, 5) + (DevLag_k = 5) \\ &\quad + (DevLag_k = 6) + (AYear_i = 1997), \\ \nu &= 2^+, \\ logit(\tau) &= poly(Adj.DevLag_k, 2)\Big). \end{aligned} \tag{ZITF 2.5}$$

The residuals in Figure 4.11(a) are scattered around 0 almost arbitrary. Only some outliers at higher development lags occur. Since focus lies on the first development lags we accept them. The Q-Q normal plot looks much better than for the last model (ZITF 2.4). It still does not look perfect, which is mainly to due to restriction of the degrees of freedom. Improvements can also be seen when calculating the SBC

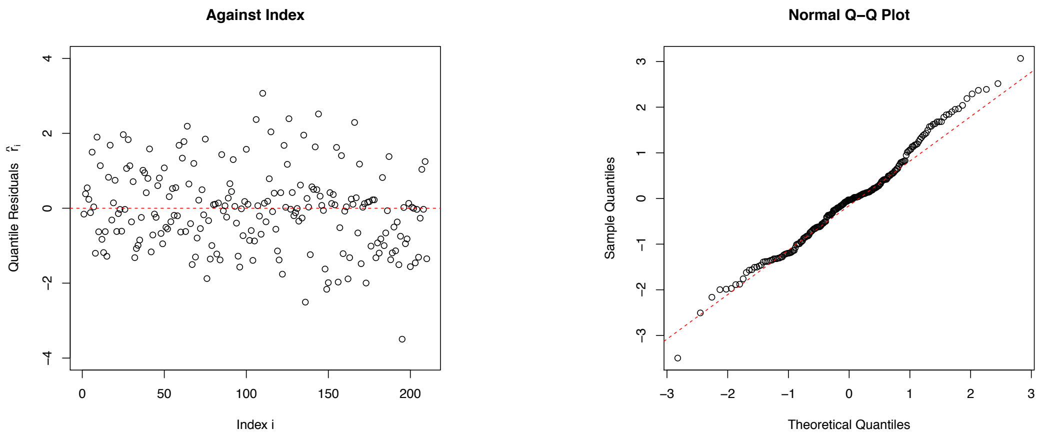
```
> ptp.gamlss.zitf2.5$sbc

[1] -1279.507
```

which decreased by

```
> ptp.gamlss.zitf2.5$sbc - ptp.gamlss.zitf2.4$sbc

[1] -49.17096
```



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure 4.11.: Both plots show a good fit for ZITF 2.5

In total we needed

```
> ptp.gamlss.zitf2.5$df.fit
```

```
[1] 20
```

degrees of freedom to model this line of business.

4.3. Results

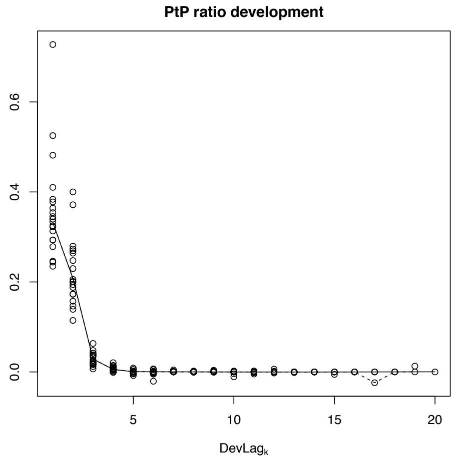
Model ZITF 2.5 is the final model for this line of business since residuals show no pattern, the Q-Q normal plot shows no huge lack of fit and the SBC is smallest among all SBC's calculated in this section for this model. To interpret the model we start with explaining the model for each parameter separately. We have used a zero-inflated t-distribution which has four parameters:

μ: The final model for μ is

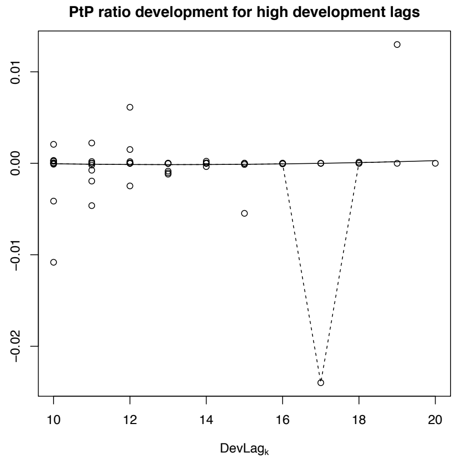
$$\mu \sim poly(DevLag_k 2) + (DevLag_k = 1) + (DevLag_k = 2) + (DevLag_k = 3) \\ + (DevLag_k = 4) + (DevLag_k = 17 \,\&\, AYear_i = 1993).$$

That means it is enough to estimate PtP ratios using a quadratic trend on the development lag and account for the first four development lags by using factors on these lags. The term for accident year 1993 at lag 17 is necessary to model the very unusual PtP ratio there.

In Figure 4.12 we see that estimates for μ are in line to what we have seen at the beginning of this section. μ is related to the mean in the way that $\mathbb{E}\left[Y\right] = (1-\tau)\mu$.



(a) Estimated μ's (solid line) describe development of PtP ratios well quad



(b) For higher lags the fit looks well. The dashed line shows the fit for accident year 1993

Figure 4.12.: Estimated values for μ for all accident years but 1993 (solid line) and for accident year 1993 (dashed line) fit well

Thus neglecting τ, a decreasing μ for increasing development lags means that we expect the PtP ratios to decrease. It can further be seen that the huge outlier for accident year 1993 at lag 17, caused by a large negative payment, is modeled by the factor.

σ**:** The final model for σ is

$$\begin{aligned}\log(\sigma) \sim\ & poly(DevLag_k, 5) + (DevLag_k = 5) \\ & + (DevLag_k = 6) + (AYear_i = 1997).\end{aligned}$$

For σ a polynomial of degree 5 was necessary plus factors for development lags 5 and 6 a factor for accident year 1997. Recall that we used a log link for σ to ensure positive values for σ and the R function `poly` fits orthogonal polynomials.

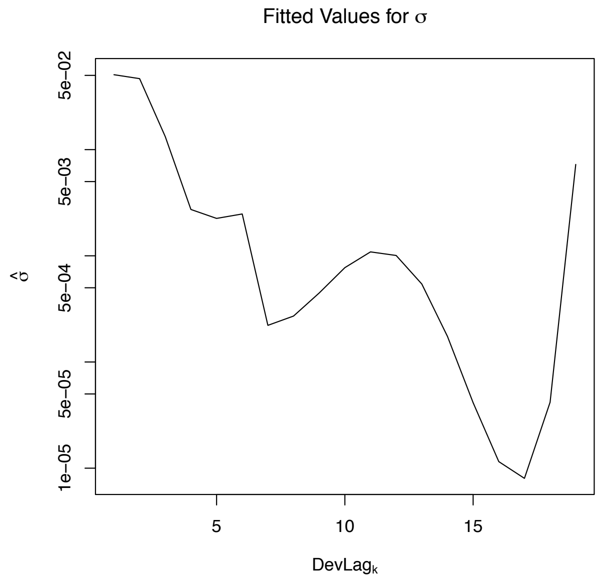


Figure 4.13.: Plot of fitted values for σ on log-scale

Figure 4.13 shows a plot of fitted values for σ for all accident years but 1997 whose slope is shifted upwards because of the factor. Fitted values are plotted on a log scale to ease identifying the slope. A decreasing trend is observable with an 'edge' for development lag 6 and an increasing trend for very high development lags. This is possibly caused by the few observations for those lags.

ν: ν is fixed at 2.1 for all observations. This is the smallest degree of freedom we could use to ensure existence of the variance.

τ: τ is model by

$$logit(\tau) \sim poly(Adj.DevLag_k, 2),$$

a polynomial of degree on the adjusted development lag, defined in (4.2).

A plot of $\hat{\tau}$ against development lag is given in Figure 4.14. Fitted values $\hat{\tau}_k$ match the observed probabilities z_k well. For development lags less than 9 we exclude the possibility of being zero by having $\tau = 0$. The quadratic curve used afterward has its maximum at lag 16 with a probability for the PtP ratio being 0 of 60%. Instead of using the adjusted development lag one could have used the original development lag well. This would then have led to probabilities greater than zero also for development lags less than 9 and smoother curve.

Recalling that neither mean nor variance of the zero-inflated t-distribution are functions of only one parameter we can't make statements about them by solely analyzing parameters

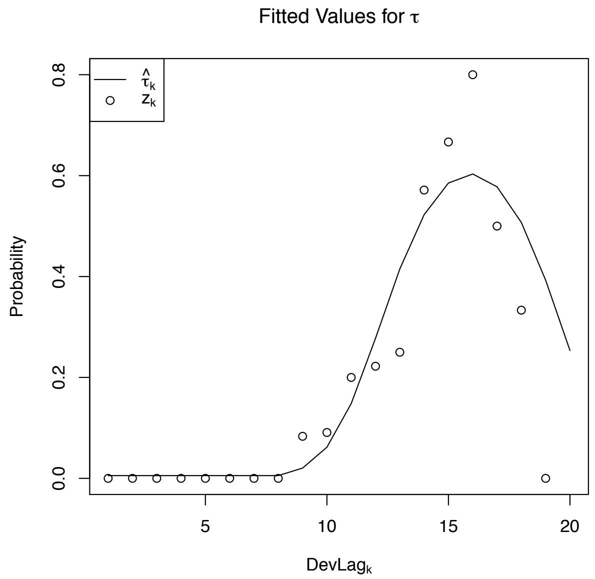


Figure 4.14.: Fitted values $\hat{\tau}_k$ and observed probabilities z_k

individually. Estimated mean and variance for the zero-inflated t-distribution are given by

$$\hat{m}_{i,k} = (1 - \hat{\tau}_{i,k})\hat{\mu}_{i,k}$$

and

$$\hat{v}_{i,k} = (1 - \hat{\tau}_{i,k}) \left(\hat{\sigma}^2_{i,k} \frac{\hat{\nu}_{i,k}}{\hat{\nu}_{i,k} - 2} + \hat{\mu}^2_{i,k} \right) - (1 - \hat{\tau}_{i,k})^2 \hat{\mu}^2_{i,k},$$

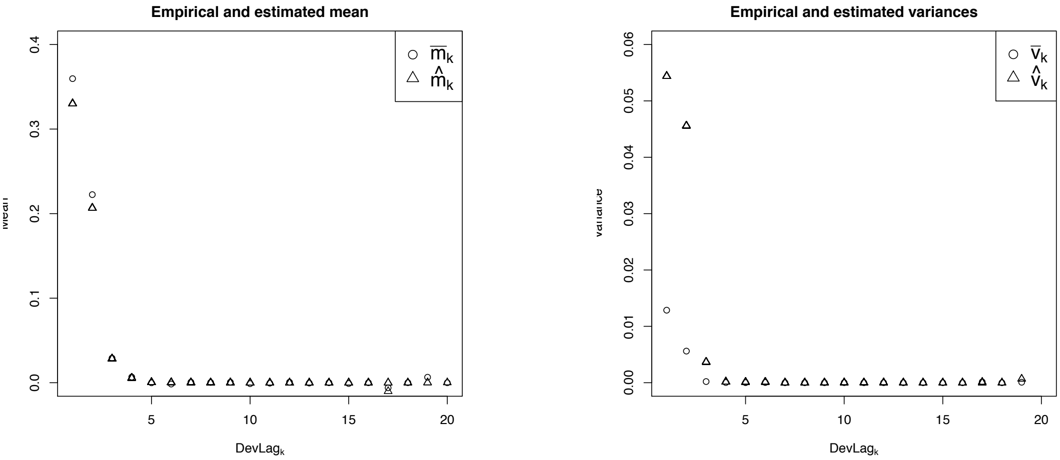
compare Section 3.3.1. Since, beside from two factors, neither mean nor variance depend on the accident year, we drop the index i. We can then compare estimated means and variances per development lag to the empirical versions, given by

$$\overline{m}_k = \frac{1}{n+1-k} \sum_{j=1}^{n+1-k} PtP_{j,k}, \quad k = 1, \ldots, n-1.$$

and

$$\overline{v}^2_k = \frac{1}{n-k} \sum_{j=1}^{n+1-k} \left(PtP_{j,k} - \overline{m}_k\right)^2, \quad k = 1, \ldots, n-1.$$

Figure 4.15(a) shows a good estimation of the mean in each development lag. Variances for the first three development lags are estimated much higher than they should as Figure 4.15(b) suggests. Since we didn't use factors for the first development lags like



(a) Empirical means $\overline{m}_k$ and estimated means $\hat{m}_k$

(b) Empirical variances $\overline{v}_k$ and estimated variances $\hat{v}_k$

Figure 4.15.: Empirical and estimated means and variances for model ZITF 2.5

we did for the mean, this could be caused by the polynomial which might not able to capture the development well.

This ends the chapter about model fitting. We have shown how to develop a model for a certain line of business and determine successively better models. The final model is still no perfect model but we stress that this is mainly driven by our restriction for ν. We also showed that detailed analyzing and understanding of the data is necessary to find a good model. Thus it takes much more time and effort to estimate a model than using e.g. the chain ladder method or a generalized linear model. The next chapter shows how limited chain ladder method and GLMs in certain scenarios are and why the GAMLSS is the better model.

At the end of this section we show an overview of all models fit for this line of business.

Model	Distr.	Parameter	Polynomials	Factors	No. Parameters	SBC
ZIG 1	ZIG	μ	poly(DevLag,5)		8	-412.1369
		$\ln(\sigma)$	1			
		ν	1			
ZITF 1	ZITF	μ	poly(DevLag,5)		9	-737.5843
		$\ln(\sigma)$	1			
		$\log(\nu)$	1			
		τ	1			
ZITF 2.1	ZITF	μ	poly(DevLag,5)		8	-674.4351
		$\ln(\sigma)$	1			
		$\log(\nu)$	2.1 (fixed)			
		τ	1			
ZITF 2.2	ZITF	μ	poly(DevLag,5)		13	-1032.481
		$\ln(\sigma)$	poly(DevLag,5)			
		$\log(\nu)$	2.1 (fixed)			
		τ	1			
ZITF 2.3	ZITF	μ	DevLag	DevLag 1,2,3,4	13	-1176.032
		$\ln(\sigma)$	poly(DevLag,5)			
		$\log(\nu)$	2.1 (fixed)			
		τ	1			
ZITF 2.4	ZITF	μ	DevLag	DevLag 1,2,3,4	15	-1230.336
		$\ln(\sigma)$	poly(DevLag,5)			
		$\log(\nu)$	2.1 (fixed)			
		$\text{logit}(\tau)$	poly(Adj.DevLag,2)			
ZITF 2.5	ZITF	μ	poly(DevLag,2)	DevLag 1,2,3,4,	20	-1279.507
				DevLag 17 & AYear 1997		
		$\ln(\sigma)$	poly(DevLag,5)	DevLag 5,6, AYear 1997		
		$\log(\nu)$	2.1 (fixed)			
		$\text{logit}(\tau)$	poly(Adj.DevLag,2)			

Table 4.3.: All models for LoB 6

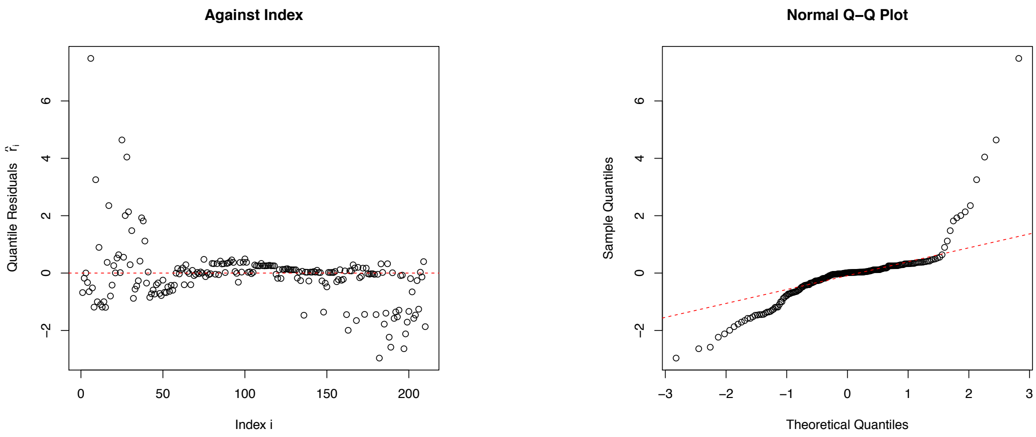


Figure 4.16.: ZIG 1

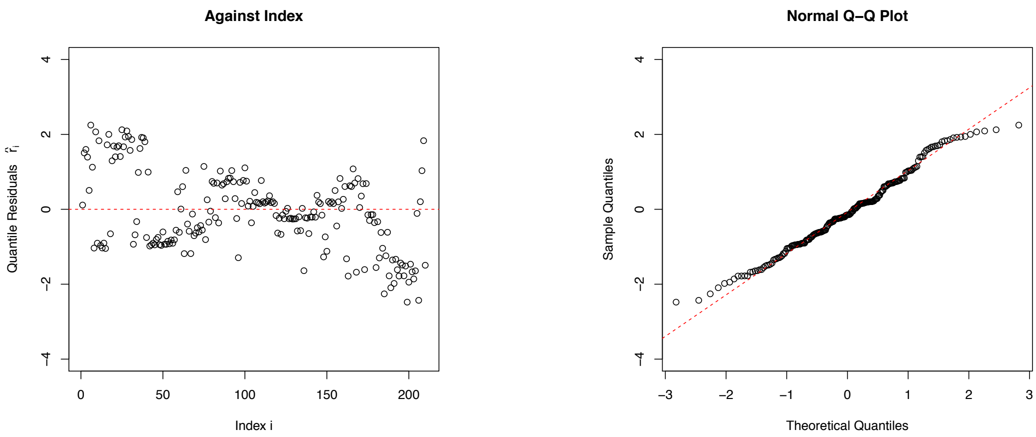


Figure 4.17.: ZITF 1

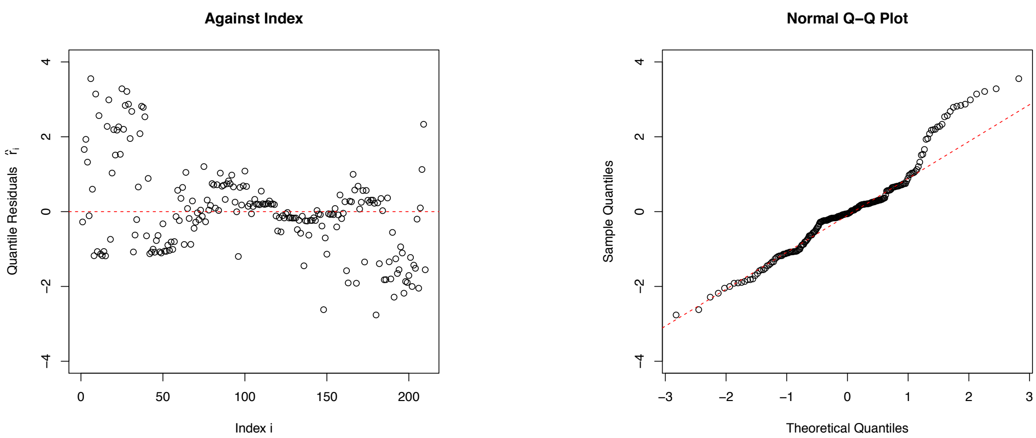


Figure 4.18.: ZITF 2.1

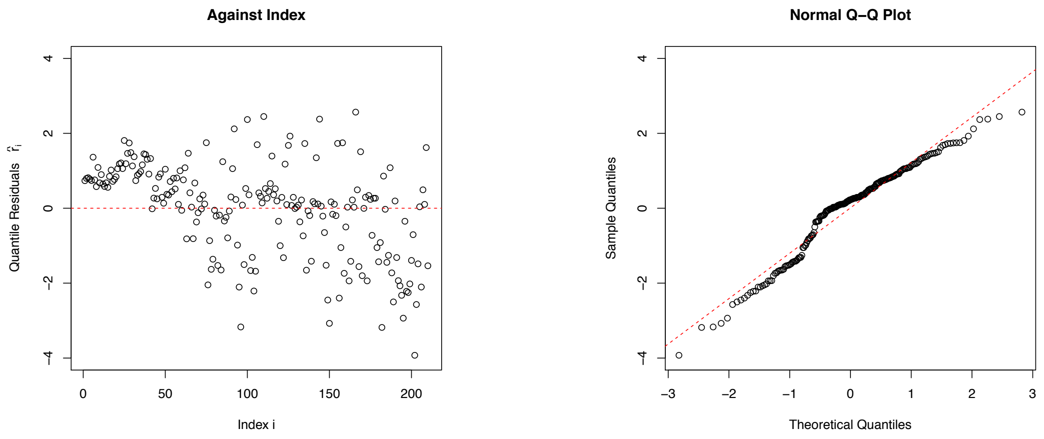


Figure 4.19.: ZITF 2.2

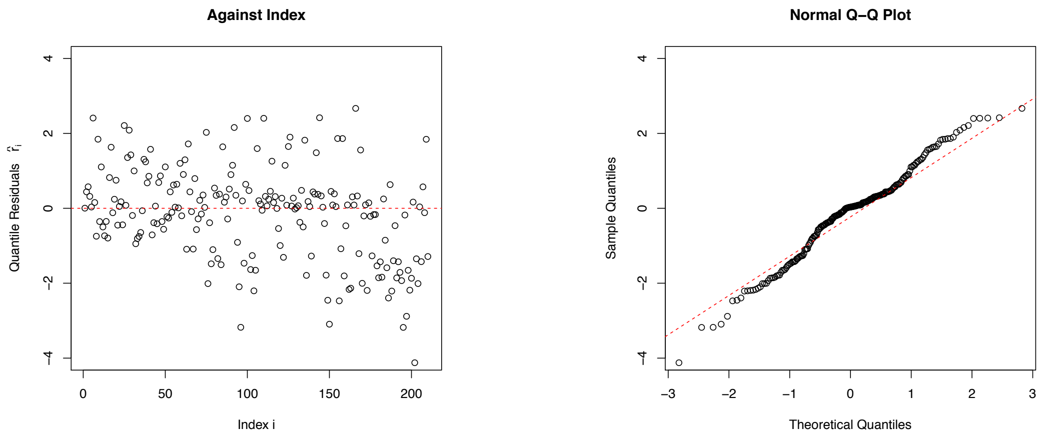


Figure 4.20.: ZITF 2.3

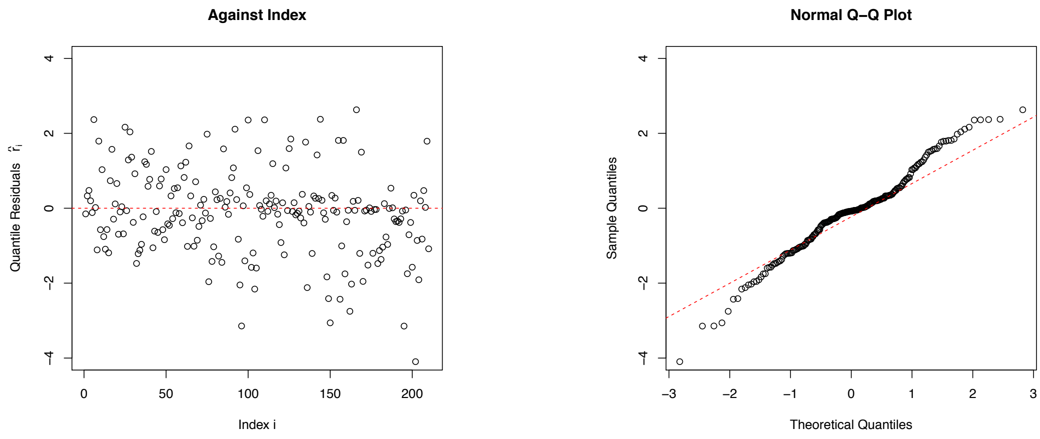


Figure 4.21.: ZITF 2.4

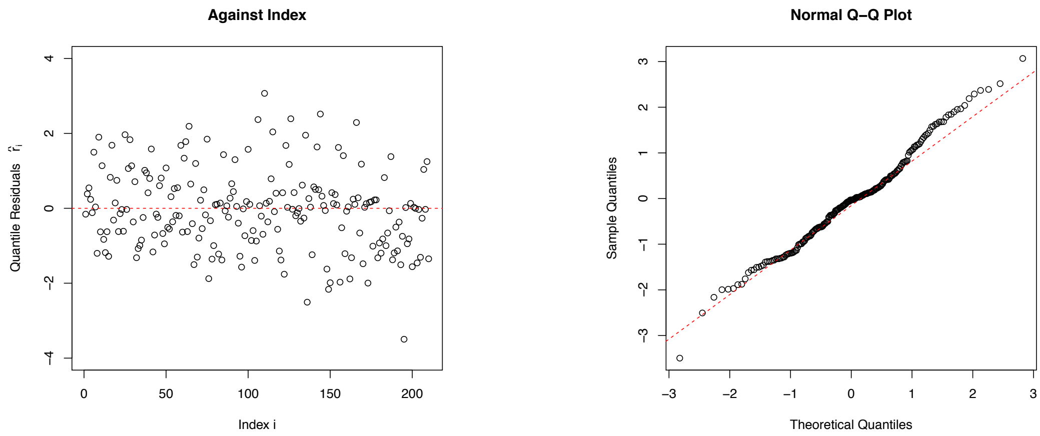


Figure 4.22.: ZITF 2.5

5. Comparison of Models

The GAMLSS introduced in the last section is a very complex model. From a practioner's point of view one may raise the question if it is really necessary to use such a model. In this chapter we want to show with some easy examples why we think it is necessary. At the beginning we will give an example where the statistical model is superior to the deterministic chain ladder method. Then we examine why especially a GAMLSS is more suitable than a generalized linear model, which has been used by Shi and Frees (2011). We recall that the chain ladder method needs (cumulative) paid losses, whereas the generalized linear model used by Shi and Frees (2011) and the GAMLSS we use need paid-to-premium ratios.

5.1. Accident Year Effects

Consider the following first two columns of a loss triangle given as in Table 5.1. Because we focus on these two columns and do not use other columns of the complete loss triangle, we omitted these.

Accident Year i	Development Lag k 1	2
1	50	115
2	60	127
3	65	140
4	70	150
5	75	170
6	80	175
7	80	170
8	70	140
9	70	130
10	65	-

(a) Cumulative losses $C_{i,k}$

Accident Year i	Development Lag k 1	2
1	50	65
2	60	67
3	65	75
4	70	80
5	75	95
6	80	95
7	80	90
8	70	70
9	70	60
10	65	-

(b) Incremental losses $S_{i,k}$

Table 5.1.: First two columns of a loss triangle

Furthermore assume the premiums are given as in Table 5.2.

Accident Year i	Premium
1	310
2	330
3	350
4	375
5	390
6	430
7	460
8	430
9	450
10	440

Table 5.2.: Premiums P_i

Then the age-to-age factors $F_{i,1}$, $i = 1, \ldots, 9$, and Paid-to-Premium ratios $PtP_{i,1}$, $i = 1, \ldots, 10$, $PtP_{i,2}$, $i = 1, \ldots, 9$, can be calculated as in Table 5.3.

Accident Year i	Age-to-Age Factor 1-2
1	2.300
2	2.117
3	2.154
4	2.143
5	2.267
6	2.188
7	2.125
8	2.000
9	1.857
10	-

(a) Age-to-Age factors $F_{i,1}$

Accident	Development Lag k	
Year i	1	2
1	0.161	0.210
2	0.182	0.203
3	0.185	0.214
4	0.186	0.213
5	0.192	0.244
6	0.186	0.221
7	0.174	0.196
8	0.163	0.163
9	0.156	0.133
10	0.148	-

(b) PtP ratios $PtP_{i,k}$

Table 5.3.: Age-to-Age factors $F_{i,1}$ and PtP ratios $PtP_{i,k}$

Figure 5.1 shows a non-linear trend for the age-to-age factors (used for the chain ladder method) and the Paid-to-Premium ratios (used for the GAMLSS) at the first development lag among the accident years. We could perform a very small analysis for the accident year 9. We store the data in a data frame and estimate factors in R. An excerpt of the first two rows of the data frame is shown below.

```
> acc.ex[1:2,]
```

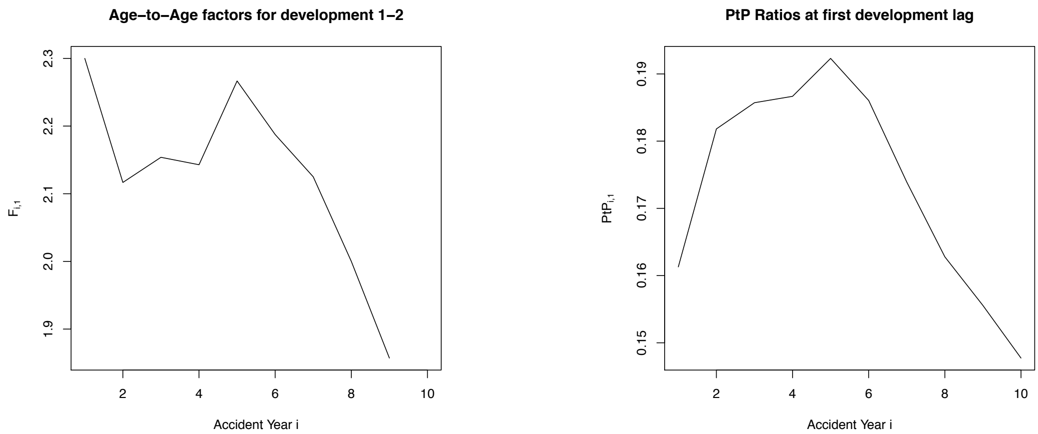


Figure 5.1.: Strong non-linear trend in both data sets

```
  Premiums Cum.Paid.Lag1 Cum.Paid.Lag2 AtA.Factor.1.2 PtP.Ratio.Lag1
1      310            50           115          2.300          0.161
2      330            60           127          2.117          0.182
  PtP.Ratio.Lag2
1          0.210
2          0.203
```

The chain ladder factor using only the first 8 accident years is

```
> round(sum(acc.ex$Cum.Paid.Lag2[1:8])/sum(acc.ex$Cum.Paid.Lag1[1:8]),3)
```

```
[1] 2.158
```

and thus $2.158/1.857 - 1 \approx 16\%$ higher than the observed one. Even excluding several accident years by using a volume weighted average of say the last 5 years still leads to an overestimation of the observed value. The GAMLSS is more flexible and allows explicit modeling of effects like these. Figure 5.1 shows an approximately quadratic trend with negative slope. Hence including a polynomial on the accident year of degree 2 as covariate will lead to much better results. Depending on the behavior in the following development lags one could use an overall polynomial (of degree 2 or maybe even more). Or if this non-linear trend appears only for the first lag one could use and interaction of polynomial and development lag 1. Performing a simple linear regression with a polynomial on accident year `AYear` of degree 2

```
> AYear <- c(1:8)
> acc.lm <- lm(PtP.Ratio.Lag1[c(1:8)]~poly(AYear,2, raw=TRUE),
+                           data=acc.ex)
```

leads to an estimate of

```
> acc.lm$coefficients %*% c(1,9,81)
```

```
          [,1]
[1,] 0.1431964
```

and thus a relative difference of $0.143/0.156 - 1 \approx -8.2\%$. Note that a linear model is a special case of the GAMLSS and thus can be fitted in R either using a standard `lm`-fit or a `gamlss`-fit.

There are several more scenarios which can cause problems to the chain ladder method. Beside a trend, unusual accident years cannot be represented properly. Caused for instance by a catastrophe one accident year could behave very different from the other ones. While the volume weighted average will smooth the development, the GAMLSS allows explicit modeling of such an accident year using a factor for this accident year. Additionally the GAMLSS can not only model the average level (i.e. measured by mean) but also a different variation (i.e. measure by variance). Suppose a line of business with first two columns like in Table 5.4. Again we omitted other columns because we don't consider them here.

Accident Year i	Development Lag k 1	2
1	55	115
2	60	125
3	65	130
4	70	155
5	75	140
6	80	185
7	80	145
8	70	164
9	70	120
10	65	-

(a) Cumulative losses $C_{i,k}$

Accident Year i	Development Lag k 1	2
1	55	60
2	60	65
3	65	65
4	70	85
5	75	65
6	80	105
7	80	65
8	70	95
9	70	50
10	65	-

(b) Incremental losses $S_{i,k}$

Table 5.4.: First two columns of a loss triangle

Furthermore assume the premiums are given as in Table 5.5.

Accident Year i	Premium
1	320
2	345
3	390
4	390
5	480
6	440
7	533
8	375
9	490
10	340

Table 5.5.: Premiums P_i

Then the age-to-age factors $F_{i,1}$, $i = 1, \ldots, 9$, and Paid-to-Premium ratios $PtP_{i,1}$, $i = 1, \ldots, 10$, $PtP_{i,2}$, $i = 1, \ldots, 9$, can be calculated as in Table 5.6.

Accident Year i	Age-to-Age Factor 1-2
1	2.091
2	2.083
3	2.000
4	2.214
5	1.867
6	2.312
7	1.812
8	2.357
9	1.714
10	-

(a) Age-to-Age factors $F_{i,1}$

Accident	Development Lag k	
Year i	1	2
1	0.172	0.188
2	0.174	0.188
3	0.167	0.167
4	0.179	0.218
5	0.156	0.135
6	0.182	0.239
7	0.150	0.122
8	0.187	0.253
9	0.143	0.102
10	0.197	-

(b) PtP ratios $PtP_{i,k}$

Table 5.6.: Age-to-Age factors $F_{i,1}$ and PtP ratios $PtP_{i,k}$, $k = 1, 2$

Figure 5.2 shows that both development factors and PtP ratios stay at the same level but their variation increases with increasing accident years. For accident year 9 the chain ladder method based upon the first eight years estimates a development factor of

```
> round(sum(acc.ex.2[1:8,3]) / sum(acc.ex.2[1:8,2]),3)
```

```
[1] 2.09
```

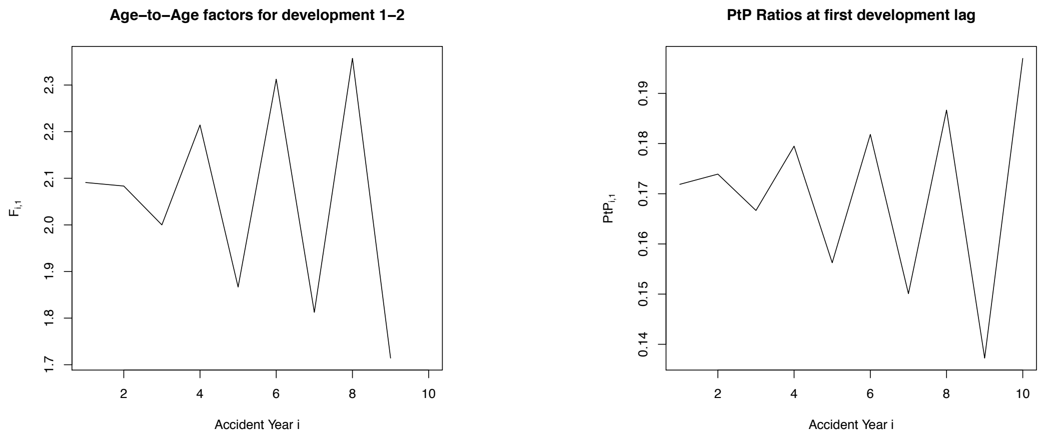


Figure 5.2.: Variation increases with increasing accident years

and hence overestimates the observed value by approx. 22%. For the GAMLSS we have the choice of using the accident year as covariate for the mean as we did in the first example. But we can see that the average PtP ratio stays the same (around 0.17), meaning that there is no trend observable. However, the GAMLSS allows explicit modeling of the scale parameter. Hence we fit a GAMLSS with a Gaussian distribution with constant location parameter and linear trend in the scale parameter, called σ:

$$PtP_{i,k} \sim \mathcal{N}\left(\mu = 1,\, \log(\sigma) = AYear_i\right)$$

In this case σ models the standard deviation and we use a log link for σ.

```
> acc.ex.2.GAMLSS <- data.frame(cbind("AYear"=c(1:8),
+ "PtP.Ratio.Lag1"=acc.ex.2[1:8,5]))

> acc.gamlss.2 <- gamlss(PtP.Ratio.Lag1~1, sigma.formula=~AYear, family=NO,
+                          control=gamlss.control(trace=F), data=acc.ex.2.GAMLSS)
```

Predicting the ratio at accident year 9 yields

```
> round(predict(acc.gamlss.2, newdata=9),3)

[1] 0.172
```

which is approx. $0.172/0.143 - 1 \approx 20\%$ higher than the observed value. What appears to be only a small improvement compared to the chain ladder method turns out to be a big one when using the fact the the GAMLSS is a statistical model. It specifies a complete distribution for each entry, so in this case we receive a Gaussian distribution for each entry. By including a linear trend for the second parameter, i.e. the standard deviation, later accident years can have a different standard deviation (and hence variance) than earlier ones. The estimated parameters for the σ-model are

```
> acc.gamlss.2$sigma.coef

(Intercept)       AYear
 -6.9097326   0.4508798
```

The estimated parameter for accident year has a positive sign, meaning that later accident years will have a larger standard deviation than earlier accident years, see Figure 5.3.

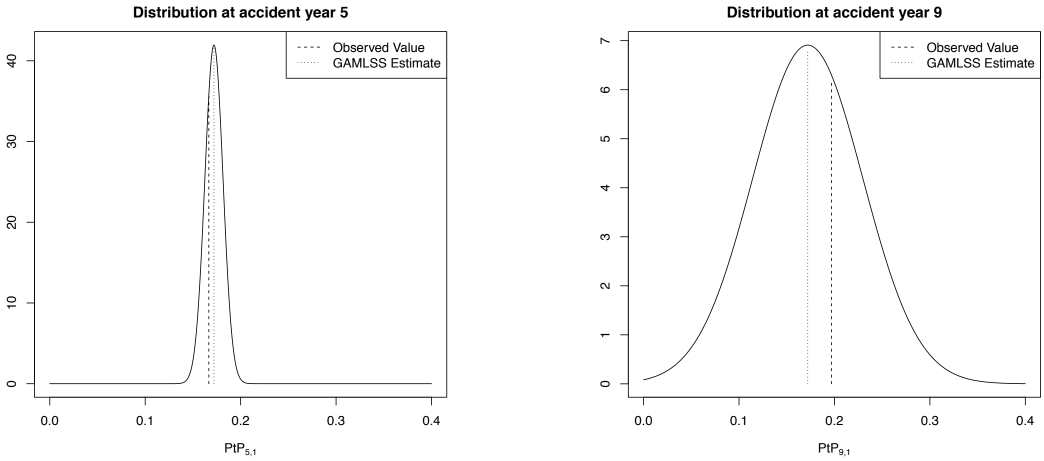


Figure 5.3.: The Gaussian distribution has a much standard deviation for accident year 9 than for accident year 5

The figure shows the huge difference between the classic chain ladder method and the GAMLSS approach. While the chain ladder method provides only a point estimate, the GAMLSS provides a distribution and hence a possibility to measure uncertainty and risk. The Gaussian distribution at accident year 9 shows much higher standard deviation than at accident year 5. Because of that any estimate between say 0.1 and 0.25 seems to be reasonable for accident year 9 while for accident year 5 the interval is much smaller. When looking at the area close to 0 for accident year 9 we can also see some support there and the probability for a negative PtP ratio is

```
> pNO(0, mu9, sigma9)

[1] 0.001443618
```

While a negative PtP ratio clearly makes no sense for the first development lag, it does for the later ones and offers the possibility of including refunds into the model.

5.2. Calendar Year Effects

Beside estimates of ultimate losses, insurance companies are often interested in estimates of next year's cash flow. The cash flow can be estimated using the chain ladder or GAMLSS method. We define next calendar year's cash flow as follows.

Accident Year i	Development Lag k 1	2	3	4	5
1	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	$S_{1,4}$	$S_{1,5}$
2	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$	$S_{2,4}$	$\mathbf{S_{2,5}}$
3	$S_{3,1}$	$S_{3,2}$	$S_{3,3}$	$\mathbf{S_{3,4}}$	
4	$S_{4,1}$	$S_{4,2}$	$\mathbf{S_{4,3}}$		
5	$S_{5,1}$	$\mathbf{S_{5,2}}$			

Table 5.7.: Cash Flow for calendar year 6 (bold incremental paid losses)

Definition 5.1: *Let $\mathcal{S}$ be a loss triangle with incremental payments. Then the* ***cash flow*** *for calendar year $n+1$ is defined as*

$$S_{n+1} := S_{2,n} + S_{3,n-1} + \ldots + S_{n,2} = \sum_{k=2}^{n} S_{k,n+2-k}.$$

$S_{1,n+1}$ is not part of the cash flow since the chain ladder method assumes all accident years to be fully developed by development lag n. An illustration of the cash flow can be found in Table 5.7.
Having a long data history there may arise problems of distortions caused by inflation. To examine influence of inflation on the estimates, we consider two data sets: one has been manually deflated while the other one has not. As a measure of inflation we used the consumer price index (CPI). We then applied the chain ladder method and the GAMLSS to both data sets without the latest diagonal, i.e. the latest calendar year. Details about the data and the GAMLSS used for this comparison can be found in Appendix A.1.

We estimated the next calendar year's cash flow and compared results to the observed cash flow. In Table 5.8 the relative differences between the estimated cash flows and the observed cash flows are shown.

	rel. difference (deflated data)	rel. difference (inflated data)
Chain Ladder	0.3548	0.3338
GAMLSS	-0.2877	-0.2877

Table 5.8.: GAMLSS estimates are closer to the observed ones than chain ladder estimates

For both data sets the GAMLSS estimate is closer to the observed one. As outlined in Appendix A.1, the GAMLSS is the same for both data sets. While chain ladder overestimates the cash flow by approx. 35.5% and 33.4%, the GAMLSS underestimates the cash flow by approx. 28.8%. The chain ladder method performs slightly better when applied to the data with inflation effects in this example, but the difference between the

two chain ladder estimates is rather small. Hence we continue to work with original data and leave inflation effects in the data.
Note at this point we neglected the different outcomes between overestimation and underestimation, which is of interested for insurer and will be discussed in more detail in the later chapters.

5.3. Dependencies

While the last two sections focused on a single line of business, one goal of this thesis is to examine dependencies across several lines of business in a portfolio. Also, similar results could have been achieved with a generalized linear model. To show why a GLM is not a suitable model to analyze dependencies, we consider Paid-to-Premium ratios of two lines of business. To measure dependencies we need the term of correlation.

Definition 5.2: *Let X and Y be univariate random variables with finite variances σ_X^2, σ_Y^2. Then **Pearson's rho** is defined as*

$$\rho(X,Y) = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\sigma_X}\sqrt{\sigma_Y}} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]}{\sqrt{\sigma_X}\sqrt{\sigma_Y}}.$$

Definition 5.3: *Let X and Y be univariate random variables and X' and Y' be independent copies of X and Y, respectively. Then **Kendall's tau** is defined as*

$$\begin{aligned}\tau(X,Y) &= \mathbb{P}\left[(X-X')(Y-Y')>0\right] - \mathbb{P}\left[(X-X')(Y-Y')<0\right]\\ &= \mathbb{E}\left[\operatorname{sign}\left((X-X')(Y-Y')\right)\right].\end{aligned}$$

Both measures take values in $[-1,1]$ with $\rho(X,Y) = 0 = \tau(X,Y)$ if X and Y are independent. Pearson's rho is widely used and thus often named Pearson correlation or just correlation. But Pearson's rho can only describe linear dependencies and is sensitive to outliers. For both cases a rank correlation like Kendall's tau is more meaningful. A rank correlation makes use of the ranks of observations, i.e. a permutation of the vector of observations such that for the ordered vector $x_{(1)} \leq \ldots \leq x_{(n)}$ holds. Two examples where Pearson's rho is not meaningful can be found in Section 7.4.
A scatterplot of PtP ratios for two lines of business is given in Figure 5.4. We can see that the relation between both lines of business seems to be almost linear and we could look at the empirical correlations between both lines of business.

```
> round(cor(LoB1$PtP.Ratio, LoB2$PtP.Ratio, method="pearson"),3)

[1] 0.689
```

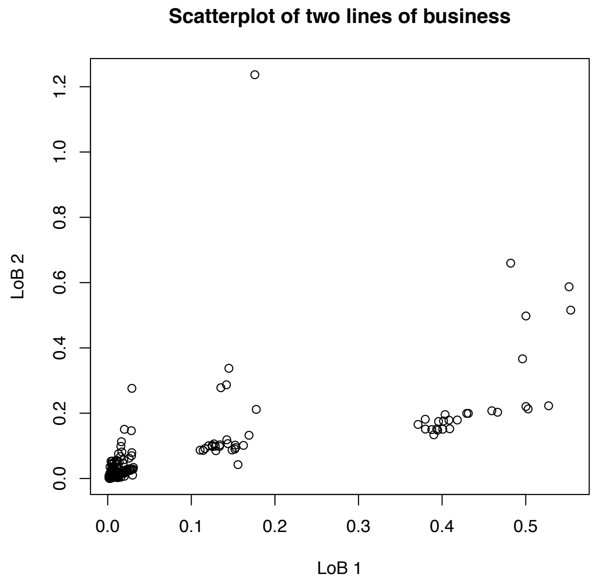


Figure 5.4.: A simple scatterplot shows a strong functional relation between both lines of business

```
> round(cor(LoB1$PtP.Ratio, LoB2$PtP.Ratio, method="kendall"),3)

[1] 0.606
```

Pearson's coefficient of correlation of 0.689 indicates moderate positive (linear) relation between both lines of business. Kendall's tau of 0.606 is slightly smaller.
But considering the scale of PtP ratios problems when using Pearson's coefficient arise. The ratios do not behave linearly for increasing development lags and it is very questionable if the PtP ratios come from a normal distribution. Both are requirements to use Pearson's coefficient in a meaningful way. Hence we do prefer rank correlation coefficients like Kendall's tau. Henceforth we only consider Kendall's tau as a measure of correlation.

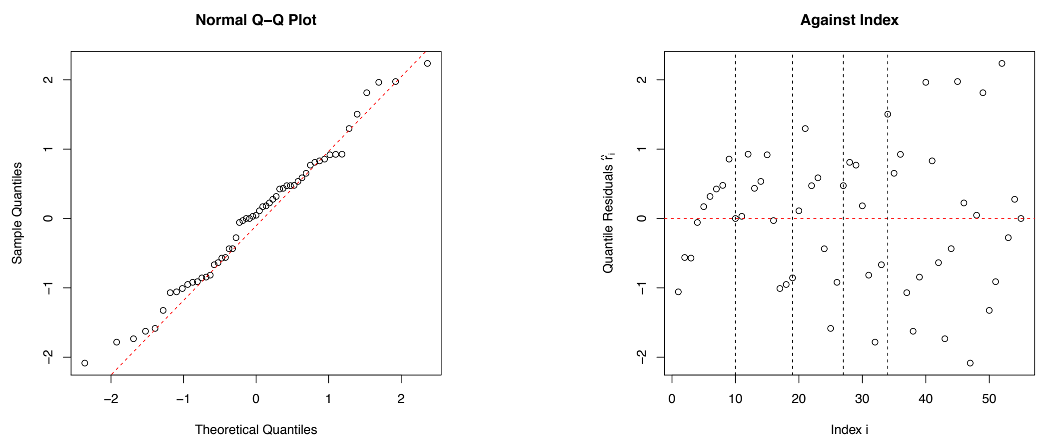
Having a correlation of roughly 0.606 still indicates strong dependence between both lines of business. The question is if it makes sense to speak of dependence between lines of business based upon this data, i.e. the PtP ratios. Ratios from both lines of business are incremental ratios. So naturally, ratios at the first development lags will be much higher than at the later ones.
While this does not affect rank correlation too much, accident year effects and calendar year effects as introduced in the last sections do. New rules by the government could lead to higher payments and thus higher ratios across all lines of business from a certain calendar year on. Hence a high correlation between two lines of business may not be due to their natural dependence only but distorted by external influences.

We therefore follow the approach presented by Shi and Frees (2011) to examine dependencies among residuals of the corresponding statistical models. They fit ordinary generalized

linear models to the data; a log-normal model and a gamma model. By doing so they argue to have modeled all dependencies caused by external effects and what remains is the true dependence between the lines of business.
Although this is an interesting idea, we think it is not enough to use GLMs as we will show in the following.

Using a generalized linear model, independent of which specific model, leads to modeling the mean (with a certain link function). Especially we assume variance homogeneity. To examine this, we use the same triangle as Shi and Frees (2011).

We fit the same to GLM to the data. PtP ratios of commercial auto are modeled by a Gamma model with accident year and development lag as factors. PtP ratios of personal auto are modeled by a log-normal model with the same covariates. The summary can be found in the Appendix A.2. We could not exactly reproduce the model for Commercial Auto, but parameter estimates are very close to the original ones. Residual plots for Personal Auto is shown below in Figure 5.5.



(a) Q-Q plot for Personal Auto model (b) Residuals against index

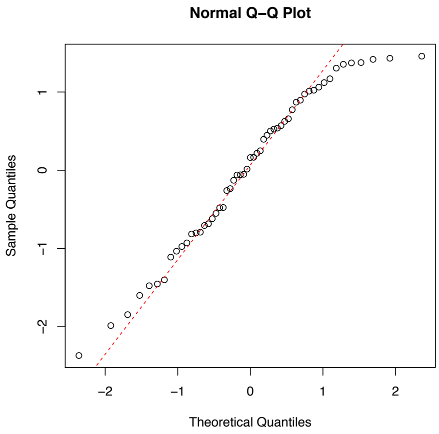
Figure 5.5.: Q-Q plot shows no lack of fit for Personal Auto GLM, but plot against the index does

The Q-Q plot for Personal Auto looks very similar to the Q-Q plot in (Shi and Frees, 2011, Figure 3). Parameter estimates are the same (compare Appendix A.2). Shi and Frees (2011) argue that it is a reasonable model. But when analyzing the residuals in a different way, things change. Figure 5.5(b) for the first development lag an upward trend for increasing accident years. For lags 2, 3 and 4 a downward trend is visible. And finally, variance increases with increasing development lags. Thus we think this is not an appropriate model and further analyses on the residuals are not meaningful.

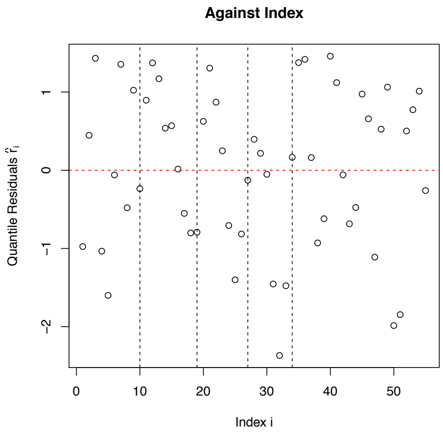
We fit a GAMLSS to the same data set. For Personal Auto the model is

$$\begin{aligned} PtP_{i,k} \sim LOGNO\big(\mu &= poly(DevLag_k, 4) + CYear_t \\ &\quad + (DevLag_k = 3) + (DevLag_k = 4), \\ \sigma &= (DevLag_k = 1)\big), \end{aligned}$$

with a log-normal distribution and identity link functions on both parameters.



(a) Q-Q normal plot of normalized quantile residuals



(b) Plot of normalized quantile residuals against index

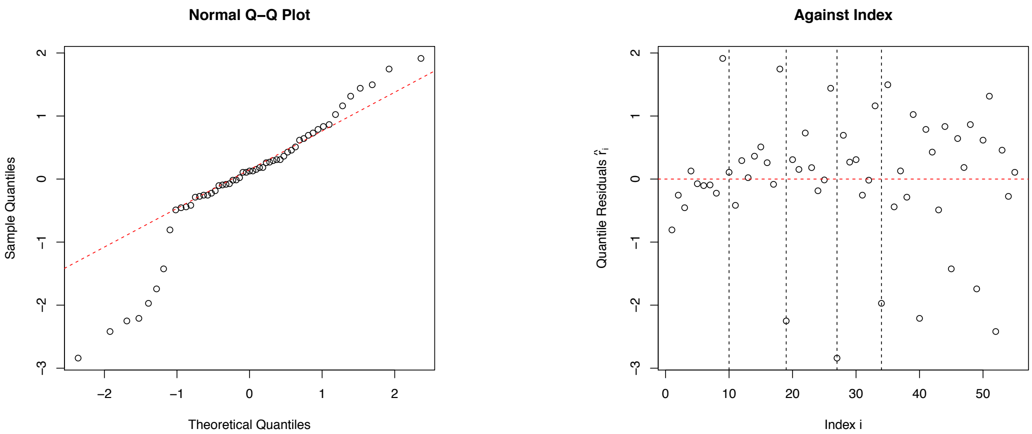
Figure 5.6.: Residual plots for Personal Auto GAMLSS

The Q-Q plot in Figure 5.6 is much smoother than for the GLM and residuals scatter randomly around zero. Both plots suggest that the GAMLSS is a better model for this line of business. Furthermore, the GAMLSS only needs 9 parameters and thus 10 less than the GLM.

The model for Commercial Auto is

$$\begin{aligned} PtP_{i,k} \sim Gamma\big(\log(\mu) &= poly(DevLag_k, 3) + (CYear_t = 1996) \\ &\quad + (CYear_t = 1997) + (DevLag_k = 1) : poly(AYear_i, 2), \\ \log(\sigma) &= CYear_t\big), \end{aligned}$$

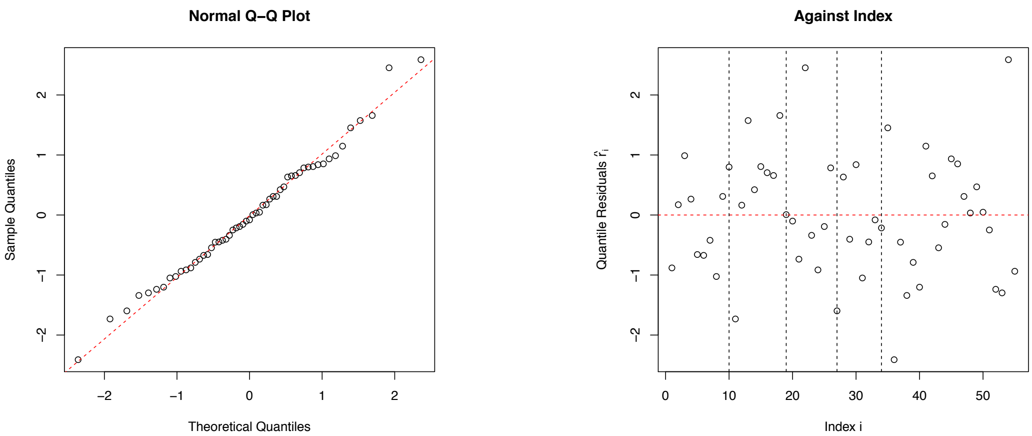
which has 12 instead of 19 parameters. Comparing the GLM with the GAMLSS in Figure 5.7 and Figure 5.8, we see the improvement in both plots. One can see a pattern for the GLM in the plot of residuals against the index. Residuals from calendar year 1996 are always very high, residuals from calendar year 1997 very low. The GAMLSS accounts for this by using a factor on these calendar years and no pattern is observable. The Q-Q plot shows a much better fit in the tails and comparing the values for the Schwarz-Bayesian criterion we see a huge drop from -175.35 to -282.87.



(a) Q-Q normal plot of normalized quantile residuals

(b) Plot of normalized quantile residuals against index

Figure 5.7.: Residual plots for Commercial Auto GLM

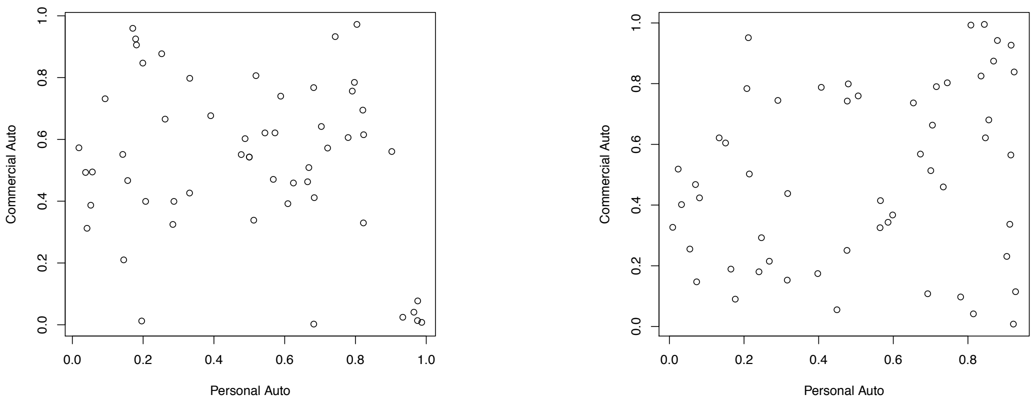


(a) Q-Q normal plot of normalized quantile residuals

(b) Plot of normalized quantile residuals against index

Figure 5.8.: Residual plots for Commercial Auto GAMLSS

So in both cases the GAMLSS is a much better model than the GLM. Even more important, the GLM shows some lack of fit and hence should not be used for further residual analyses.



(a) Residuals from GLM

(b) Residuals from GAMLSS

Figure 5.9.: Scatterplots of residuals from both approaches

For the scatterplot in Figure 5.9 note that residuals have been transformed to normalized residuals and take values in $[0, 1]$. It is difficult to see any relation between Commercial Auto and Personal Auto from these plots. Hence we calculate the correlation (Kendall's tau) of the residuals. For residuals from the two GAMLSS it is

```
> cor(pnorm(Comm.GAMLSS$resid),pnorm(Pers.GAMLSS$resid), method="kendall")

[1] 0.1609428
```

whereas the correlation of residuals from the GLMs is

```
> cor(pnorm(Comm.GLM$resid),pnorm(Pers.GLM$resid), method="kendall")

[1] -0.07814079
```

Note that the GLMs have been estimated in the GAMLSS framework but remain GLMs. The correlation between both approaches is very different. For the GLM we have a negative correlation (which is in line with Shi and Frees (2011)) while there is a positive correlation for the GAMLSS. Hence interpretation of results strongly depends on the model and will be very different. Considering residuals from the GLM one could argue that a small diversification effect is present and hence the portfolio reserve could be smaller than the sum of individual reserves. However, the GAMLSS leads to the opposite result, i.e. that the portfolio reserve should be even higher than the sum of individual reserves.
But having in mind that the GLM shows some lack of fit, the correlation from the GLM is not very meaningful. On the other side the GAMLSS for both lines of business fit well and hence also interpretation of the correlation of residuals is reasonable.

Summary

We showed that using a statistical model enables us to incorporate accident year and calendar year effects much better than the chain ladder method. Since we use Paid-to-Premium ratios for the GAMLSS, inflation effects are not present. Among statistical models we gave in insight why we think a GAMLSS is superior to standard GLM for modeling PtP ratios. It is very important to check model assumptions and analyze residuals in multiple ways. It also makes sense to use continuous covariates rather than factors to keep the number of parameters low, especially for small triangles
Therefore GAMLSS will be used in the following sections for all lines of business to ensure that the chosen models are indeed appropriate. Dependence can then be analyzed on the residuals of those models.

6. Future Claims Analysis

Goal of this chapter is to analyze the ability of the GAMLSS of predicting future claims. As a benchmark we compare results to the chain ladder method. To analyze future claims we

- restrict the data set by excluding the latest calendar year
- fit a GAMLSS to the restricted data set and apply the chain ladder method to the restricted data set
- predict the next calendar year with both methods
- compare predicted values to observed values.

This procedure is motivated by the fact that it provides a relative easy way to check the short term prediction power of both methods. And estimating PtP ratios or paid losses for the next calendar year is important for an insurer since that is an estimate for next calendar year's cash flow which can have an effect e.g. on business decisions. The cash flow for calendar year $n+1$ is defined as

$$S_{n+1} = \sum_{k=2}^{n} S_{k,n+2-k},$$

compare Definition 5.1.

Clearly, the cash flow for year $n+1$ depends on things happening in calendar year $n+1$. Unusual events in that year could result in a cash flow that is very different from the predicted one. To minimize risk of wrong inference caused by an unusual calendar year, we do not compare only one cash flow but six. Data is available until 2011 and we estimate cash flows for calendar years 2006 to 2011.
As mentioned at the beginning of the last section data until 2011 includes 231 observations. Excluding the last calendar year reduces the data set to 210 observations, on which the GAMLSS model is fitted and chain ladder factors are calculated. An overview of the amount of observations available for each cash flow estimation is shown in Table 6.1.
For the chain ladder method we only need to apply the chain ladder method to smaller triangles. For the GAMLSS it would have been eligible to do so as well. However, it turns out that more work has to be done here. One reason is that some covariates simply cannot enter a model. Suppose a model has been fit using data until 2007. Then there is no sense in using a factor for accident year 2008 for a parameter. But when using data until 2010 the factor might be needed.

Data until	Number of obs.	Predict cash flow for
2005	120	2006
2006	136	2007
2007	153	2008
2008	171	2009
2009	190	2010
2010	210	2011

Table 6.1.: Number of available observations.

We also observed some convergence problems for smaller triangles. This was probably caused by the attempt to fit too complex models to data with too few observations. Note that for the smallest triangle, i.e. data until 2005, only 120 observations are available. Thus for each line of business models for each restricted data set had to be fitted separately. Giving detailed explanations of each model would go beyond the scope of this thesis. Hence summarized results about the models can be found in the appendix.

Another point we would like to address is the question which is an appropriate estimate for unknown $S_{i,k}$ in the GAMLSS framework. For each observation $S_{i,k}$, $i, k = 1, \ldots, n$, the GAMLSS estimates a distribution. A natural suggestion could be to use the mean of the distribution. Since all observations are assumed to be independent the cash flow could be estimated as

$$\hat{S}_{n+1} := \mathbb{E}\left[S_{n+1}\right] = \mathbb{E}\left[\sum_{i=2}^{n} S_{i,n+2-i}\right] \overset{\text{ind.}}{=} \sum_{i=2}^{n} \mathbb{E}\left[S_{i,n+2-i}\right].$$

While this makes sense for symmetric distributions it is not necessarily meaningful for skewed distributions. Then other statistics like median or mode make more sense. But in contrast to the mean, e.g. the median of a sum of random variables is not the sum of medians of the single random variables. And since we have used complex distributions such as the skew exponential power type 1 distribution it would be difficult to find an analytic expression for the median.
Thus, for calendar year $n+1$, we simulate $L = 5000$ cash flows $\hat{S}^{l}_{n+1}$, $l = 1, \ldots, L$, and examine the predictive distribution. To simulate a cash flow, random number from the underlying distribution in each cell of the diagonal are drawn and summed up. For the set of 5000 cash flows empirical mean, median and quantiles are available. Depending on the shape of the distribution further analyses have to be done, separately for each line of business and each calendar year.

6.1. Line of Business 1

We start with the first line of business. The six models are shown in Table B.1.1. For all six models a SEP1 distribution has been used with constants ν and τ. For μ and σ the formulas vary between the different models.

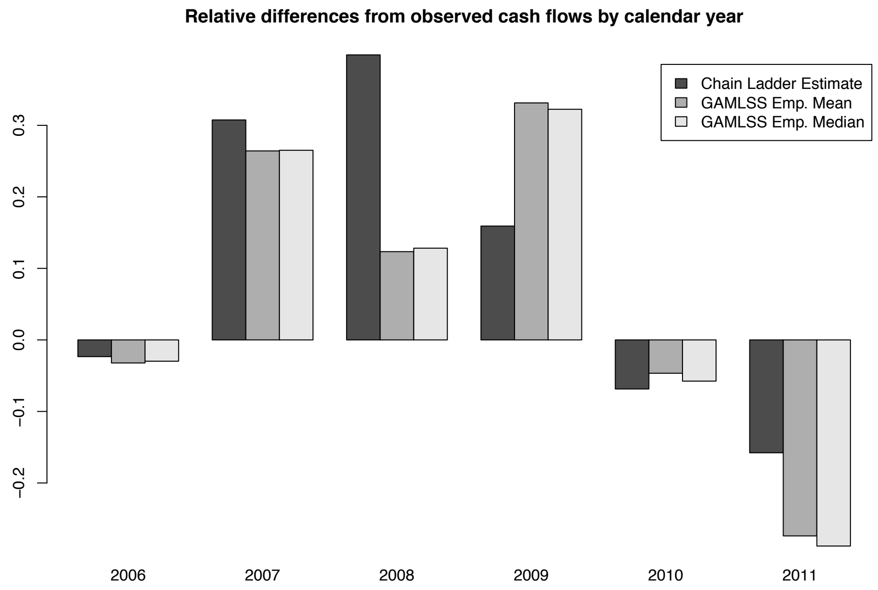


Figure 6.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 1

Figure 6.1 shows relative deviations of chain ladder estimates and empirical means and empirical medians of GAMLSS simulations from the observed cash flows. Both methods underestimate the observed cash flow three times and overestimate it three times. While for 2006, 2007 and 2010 estimates from both methods are close to each other, there are bigger differences for the other calendar years. But it is not clear from this plot what causes the differences to the observed cash flow for each method, why results differ between methods or how significant differences between estimates of both estimates are. To answer these questions Figure 6.2 shows density plots for all six estimated cash flow distributions. To maximize readability for each cash flow distribution axes vary between plots. It shows

- what shape the distribution has
- whether the mean or the median should be used for GAMLSS
- how big differences between estimated cash flows are.

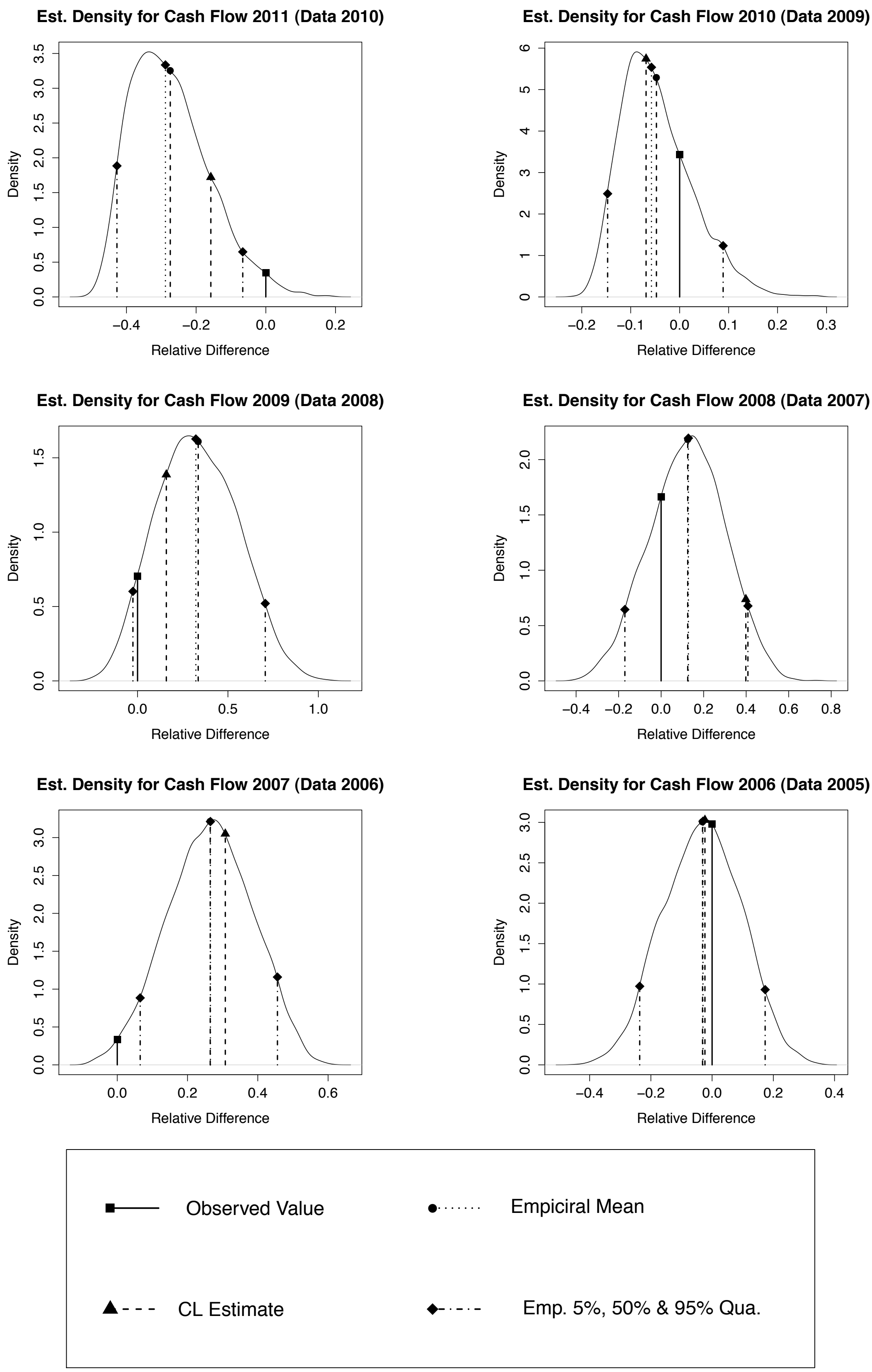


Figure 6.2.: Predictive distributions for LoB 1

There are a couple of observations that can be made for the calendar years.

2011: The chain ladder estimate is approximately 16% below the observed cash flow for calendar year 2011. Both, mean and median from the GAMLSS are much lower and underestimate the observed paid loss by approx. 30%. So both methods strongly underestimate the observed paid loss for this calendar year.
The density is right-skewed such that the median would be the more conservative, i.e. higher estimate. We see that the observed cash flow is even beyond the empirical 95%-quantile.
To find out what caused this huge differences we have to look at the underlying data set. Clearly, big differences for the first development lags have a big influence on the total difference since paid losses are much higher in absolute terms for these lags.
Checking the data shows that an unusually high PtP ratio is observable for accident year 2010 at lag 2. The development factor for this year is also higher than the chain ladder factor, but not that extreme. Thus the chain ladder method performs better in this case. But none of the methods was able to forecast such an unusual loss.
Within the GAMLSS framework, the probability that a cash flow greater or equal to the observed cash flow occurs is only 2% and hence very rare scenario.

2010: Again the density plot shows a slightly right-skewed distribution. In contrast to the first case estimates for calendar year 2010 with data until 2009 are much closer to the observed cash flow. Chain ladder estimate, GAMLSS mean and median all underestimate the observed cash flow by less than 10%. The chain ladder estimate is slightly below the GAMLSS estimate.
The underlying data has no noticeable observations for calendar year 2010, so differences arise from the methods. Note at this point that the outlier for calendar year 2011 can be neglected since we forecast only one period.

2009: In contrast to estimates for 2010 and 2011, estimates for 2009 from both methods overestimate the observed loss. The chain ladder method overestimates it by roughly 16% and the GAMLSS by 34%. We see a very symmetric distribution, hence mean and median are almost the same. No outlier could be detected in the data. Thus one should prefer the chain ladder estimate in this case.

2008: Again we see very symmetric distribution and empirical mean and median are both at 13%. Here the GAMLSS outperforms the chain ladder method whose estimate is about 40% higher than the observed cash flow. The reason for that is a very low development factor for accident year 2007 while the PtP ratio for that year is not noticeable.

2007: Mean and median of the almost symmetric distribution are both at approx. 26% which is slightly below the chain ladder estimate (31%). Both estimates strongly overestimate the observed cash flow although data gives no obvious explanation for this. PtP ratios and development factors for calendar year 2007 are slightly lower than for previous years such that the big difference is caused by the sum of all the smaller differences.

2006: Chain ladder estimate, GAMLSS mean and median are all very close to the observed cash flow. They are only 2-3% below the observed cash flow. So for this calendar year both methods yield very good estimates.

After examining the six plots no method can be preferred. Three times chain ladder estimates is closer to the observed cash flow, three times mean or median of the GAMLSS is closer. For the GAMLSS differences between the mean or median are rather small, only for the cash flows of calendar year 2010 and 2011 they are visible. While the big differences for calendar year 2011 (GAMLSS) and for calendar year 2008 (chain ladder) were caused by unusual observations, differences for other calendar years seem to arise from the methods. For chain ladder explanations are mainly of the type 'observed factor is higher/lower than estimated one'. It is not that easy for the GAMLSS. In complex models many things influence the estimation of future PtP ratios and detailed analyses are necessary to fully understand what causes the differences. There might also be a trade off between a model which yields good estimations of cash flows on other models which fit better in general, so model choice is can be difficult.

6.2. Line of Business 2, 3, 4 and 5

Analyses for all other lines of business could be done in a similar way. However, we don't present detailed results for each lines of business. We only show results for another line of business, line of business 3. Here the GAMLSS outperforms the chain ladder method and does so not by chance. Figures and models for lines of business 2, 4 and 5 can be found in Appendix B.2, B.4 and B.5.

The six models for line of business 3 can be found in the Table B.3.1.

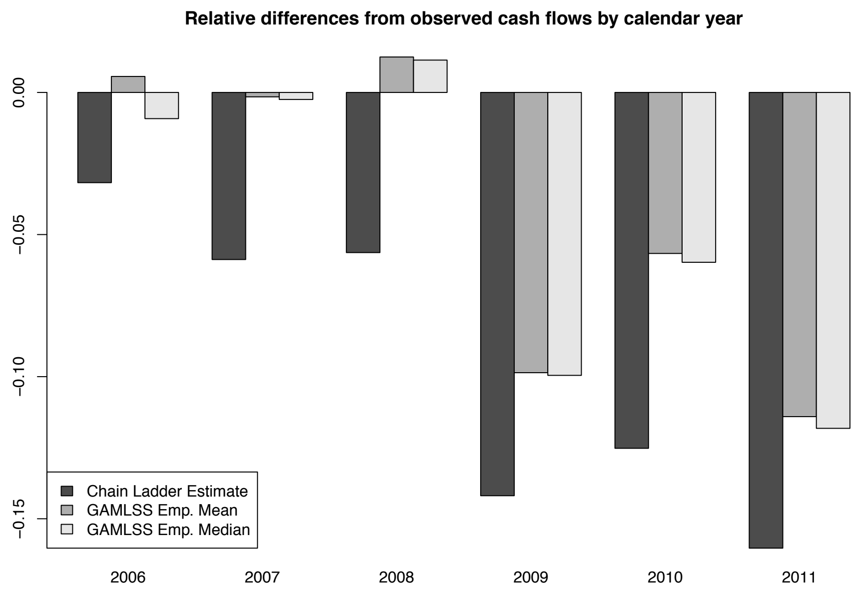
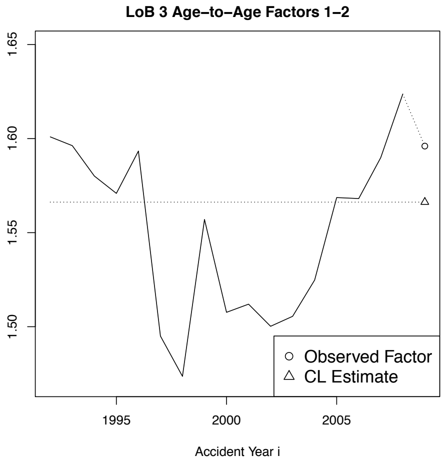
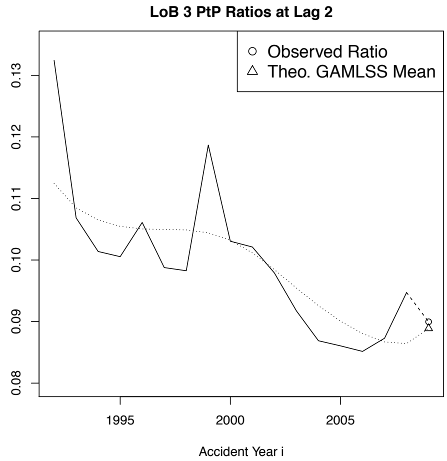


Figure 6.3.: Chain ladder estimates, empirical means and empirical medians from GAMLSS for LoB 3

Figure 6.3 shows that the GAMLSS mean and median are closer to the observed cash flows than chain ladder estimates for all calendar years. Although, in contrast to line of business 1, the scale is smaller (relative differences to not exceed -20%, the difference between GAMLSS and chain ladder method is remarkable. When looking at the data set we find a situation as described in Figure 5.1.



(a) Age-to-Age factors 1-2 for LoB 3



(b) PtP ratios of LoB 3 at development lag 2

Figure 6.4.: Observed age-to-age factors 1-2 (a) and PtP ratios for development lag 2 (b) for line of business 3

Figure 6.4(a) shows the individual age-to-age factors for each accident year from development lag 1 to lag 2 of data until 2009. The chain ladder estimate, based upon all but the latest accident year (in this data set 2009), is lower than the observed one. Note that although the relative difference between the estimated and observed factor is rather small, the relative difference between the estimated and observed incremental paid loss is much higher. When forecasting one period the relative difference between estimated incremental payment and observed incremental payment is

$$r^I_{i,k} := \frac{\hat{S}_{i,k}}{S_{i,k}} - 1 = \frac{\hat{C}_{i,k} - C_{i,k-1}}{C_{i,k} - C_{i,k-1}} - 1 = \frac{(\hat{f}^{CL}_{k-1} - 1)C_{i,k-1}}{(F_{i,k-1} - 1)C_{i,k-1}} - 1 = \frac{\hat{f}^{CL}_{k-1} - 1}{F_{i,k-1} - 1} - 1,$$

where $F_{i,k-1}$ is the observed age-to-age factor and $\hat{f}^{CL}_{k-1}$ the estimated factor from the restricted data set. This obviously is not the same as just the relative difference of the chain ladder factor and observed age-to-age factor,

$$r^F_{i,k} := \frac{\hat{f}^{CL}_{k-1}}{F_{i,k-1}} - 1.$$

Especially for $F_{i,k-1}$ and $\hat{f}^{CL}_{k-1}$ close to 1 the difference between $r^I_{i,k}$ and $r^F_{i,k}$ will be huge. Although Figure 6.4(a) shows a relative difference of age-to-age factors of only roughly $r^F_{2009,2} = -1.8\%$, the relative difference of incremental payments is $r^I_{2009,2} = -5\%$. Since age-to-age ratios are much bigger than 1 (around 1.5 - 1.6), $r^F_{2009,2}$ and $r^I_{2009,2}$ are not too different. For other lines of business with smaller development factors 1-2 or for later age-to-age factors the difference can be much higher.
We also see a functional relation between accident year and age-to-age factor which could be described as a quadratic trend. However, the chain ladder method cannot account for

this and only estimates a constant factor (dotted line). So in this case the chain ladder method cannot estimate the development factor properly.
In contrast to that the GAMLSS estimate is much closer to the observed paid loss. Figure 6.4(b) shows the observed PtP ratios at development lag 2. The estimated PtP ratio is only 1% below the observed one and hence much closer to it than the chain ladder estimate to the development factor. In contrast to chain ladder estimates the relative difference of PtP ratios is equal to the relative difference of incremental paid losses since

$$r^I_{i,k} = \frac{\hat{S}_{i,k}}{S_{i,k}} - 1 = \frac{\widehat{PtP}_{i,k} \cdot Pr_i}{PtP_{i,k} \cdot Pr_i} - 1 = \frac{\widehat{PtP}_{i,k}}{PtP_{i,k}} - 1,$$

where $PtP_{i,k}$ is the observed Paid-to-Premium ratio, $\widehat{PtP}_{i,k}$ its estimate and Pr_i the premium for accident year i. So the difference can directly be seen in Figure 6.4(b). Furthermore the dotted line shows the estimated PtP ratios for the second development lag and we see that the curve describes the actual observations well. Here a linear trend on the accident year for development lag 2 on μ and a polynomial of degree 4 on accident year for all development lags for σ has been used, compare Table B.3.1.
The same observations for chain ladder estimates and GAMLSS estimates can be made for all parts of the cash flow for this calendar year. Similarly the huge differences between chain ladder and GAMLSS estimates for the other calendar years are mainly caused by this fact. This explains why the GAMLSS outperforms the chain ladder method for this line of business so much.

Summary

We showed in this section how well chain ladder method and GAMLSS can estimate next calendar year's cash flow. We paid special attention to the point estimates of both methods. While for line of business 1 results were mixed, the GAMLSS outperformed chain ladder tremendously for line of business 3. One may raise the question why this is not the case for all lines of business. The GAMLSS is much more powerful and ideally should perform better for all lines of business. But it still is a statistical model which depends on the underlying data. We have seen for line of business 1 that a couple of outliers are in the data set. Hence it is no surprise that the GAMLSS can't model these outliers well. On the other side, if data follows a pattern like for line of business 3, we showed that a GAMLSS can lead to much better estimates than the chain ladder method. Hence no method can be preferred in all cases and detailed analyses of underlying models and data sets are required to choose the best and most reliable method.

Beside the absolute values of the relative differences it is also important to distinguish between differences with positive or negative signs. A positive sign means the cash flow was overestimated. In that case the insurer would have accounted for more losses than eventually occurred. Underestimation is much more risky and could lead to serious problems for an insurer. So one might not always prefer the method with the smallest absolute value of relative differences but the one which is more conservative, i.e. more likely to overestimate losses.

7. Reserve Estimation and Dependence Analysis

Beside estimates for next calendar year's cash flow the main goal of reserving still is to estimate reserves, i.e. *all* future obligations. As mentioned in earlier chapters we don't assume further changes of claims after the last development lag. Thus the reserve for accident year i is defined as

$$R_i := \sum_{k=n+2-i}^{n} S_{i,k}, \quad i = 2, \ldots, n.$$

Both methods, chain ladder and GAMLSS, yield estimates of unknown future payments and therefore of the reserves. Similar as for cash flow analyses the chain ladder estimates include no information about uncertainty. Hence bootstrapping is commonly used in insurance companies to obtain predictive reserve distributions and measures of uncertainty like the mean squared error of prediction. Bootstrapping has been introduced in Chapter 1. The GAMLSS on the other side provides predictive distributions for each unknown future payment but not for the sum of them. While in some cases, i.e. if all payments follow a Gaussian distribution, the sum of those random variables can easily be determined, this is not the case for distributions like the skew exponential power type 1 distribution. Since this distribution has been used a couple of times, a different technique is necessary. We used a similar approach as for the predictive cash flow distributions in the last section and simulated 2000 sets of future payments. But in contrast to the cash flow projection we incorporated parameter uncertainty. This is necessary since the chain ladder bootstrapping incorporates parameter uncertainty as well and otherwise result would not be comparable. To also incorporate parameter uncertainty, new loss triangle have been generated from random numbers of the underlying distributions. The GAMLSS have been refitted and future payments have been estimated as predictions of the GAMLSS.

Predictive distributions of reserves can then be estimated per accident year or for the whole line of business. It is often not necessary to analyze each accident year separately since the final reserve is allocated to the whole line of business. Therefore only a short analysis of reserves by accident is done on the following pages. Especially predictive distribution are only considered for the total reserve. Details about the models used for reserve estimation can be found in the appendix.

7.1. Line of Business 3

For reserve estimation a GAMLSS with SEP1-distribution has been used for this line of business. Details about the fit can be found in Appendix B.3.2. In Section 6.2 cash flow analyses showed that the chain ladder method tends to underestimate cash flows for this line of business. Hence we expect the chain ladder reserve estimates to be smaller than GAMLSS reserve estimates. Nevertheless the variation of estimates can behave very differently and strongly depends on the models used. We illustrate this in Figure 7.1, where the estimated ultimate losses per accident year are shown together with empirical quantiles, indicating variation of estimates. Beside variation of ultimate loss estimates, trends could be detected in such a plot.

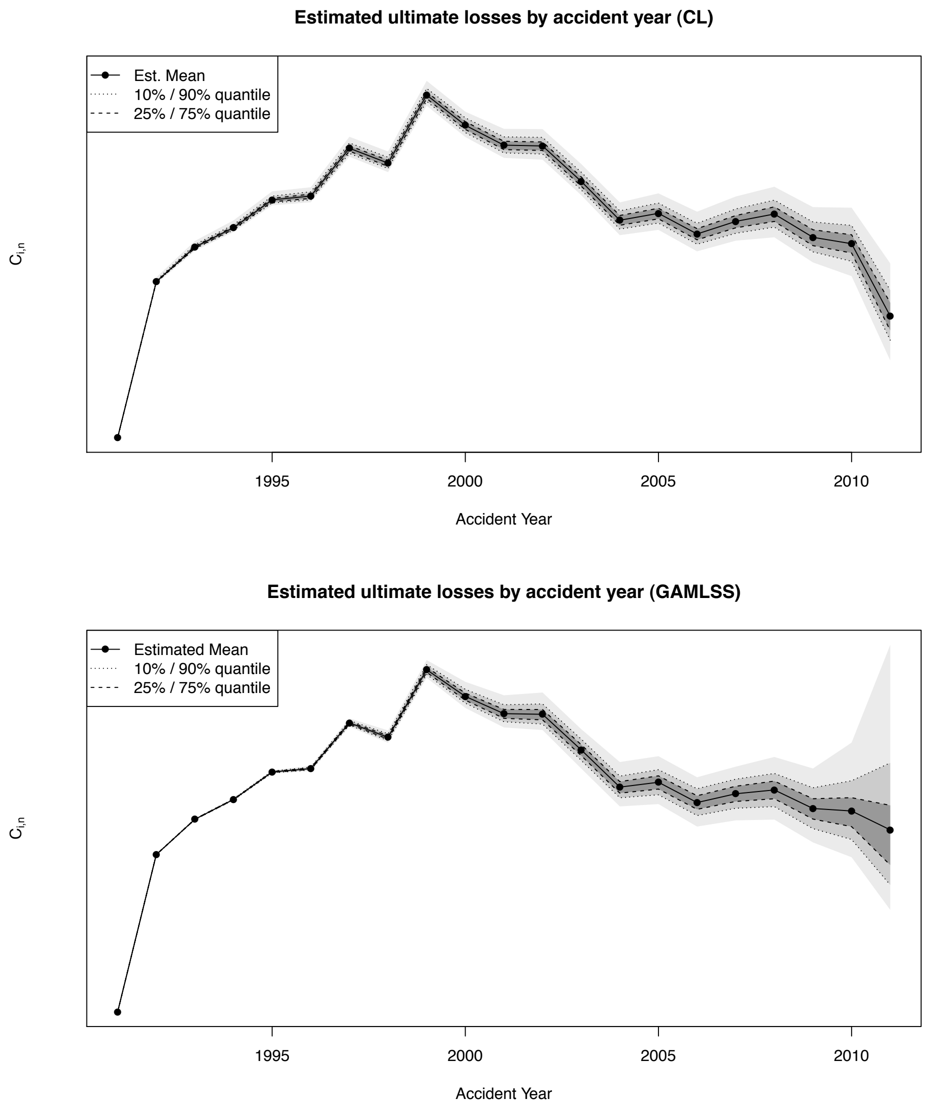


Figure 7.1.: Ultimate loss estimates for both methods (LoB 3)

Figure 7.1 shows estimated ultimate losses from both methods by accident year. Recall from Chapter 1, that ultimate losses are estimated by

$$\hat{C}_{i,n} = C_{i,n+1-i} + \hat{R}_i, \quad i = 2, \ldots, n.$$

Both plots show very similar behavior in terms of the mean, with a bigger difference only for accident year 2011, where the GAMLSS estimates a much higher ultimate loss than the chain ladder method. This is in line to what we have seen in Section 6.2. There the chain ladder method underestimated especially the first development lags, i.e. the first claims for the latest accident year(s). Because these estimates have the biggest influence on the estimated ultimate loss, this causes the difference. The bands between the 10% and 90% quantile (grey) or 25% and 75% quantile (dark grey) don't differ too much in terms of width for older accident years. But the light grey band, covering the area between the 1% and 99% quantile, is much wider for GAMLSS estimates than chain ladder estimates for recent accident years. So GAMLSS estimates for unknown future claims vary a lot more than chain ladder estimates in this case. Note that when considering reserves instead of ultimate losses the variation of estimates will be the same.

We combine the results of individual accident years by summing up all simulated unknown future payments. Here we are especially interested in the total reserve distribution and not the total ultimate loss distribution.

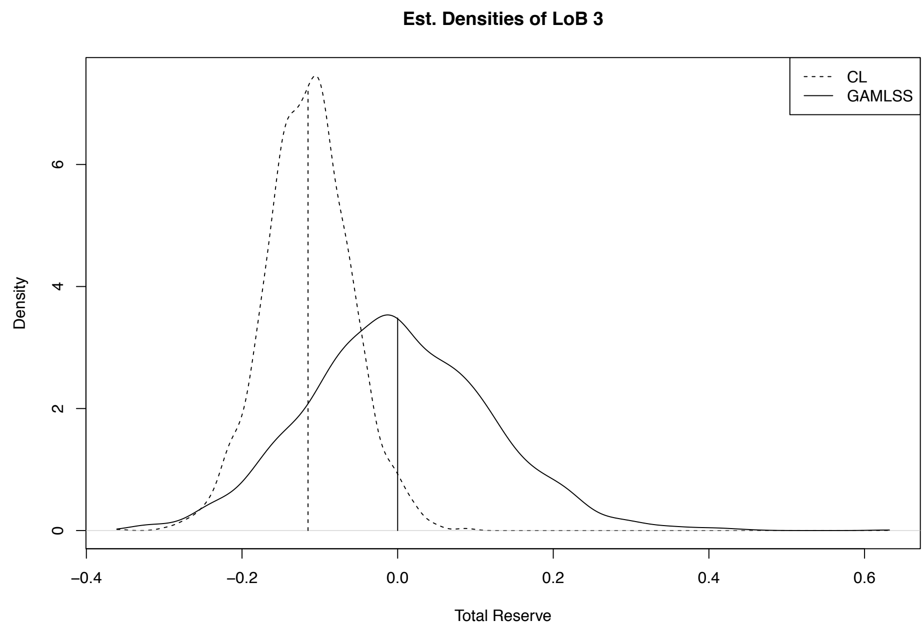


Figure 7.2.: Estimated densities for line of business 3

Figure 7.2 shows estimated densities of total reserves from chain ladder bootstrapping and GAMLSS simulation. Reserves are relative to and centered around the estimated total reserve from the GAMLSS simulation. The mean of chain ladder bootstrapping

	GAMLSS	CL
Mean	0.0000	-0.1152
Variance	0.0142	0.0030

Table 7.1.: LoB 3 Total reserve statistics

estimates is 11.52% smaller than the mean of GAMLSS estimates. Furthermore the variation around the mean is much bigger in the GAMLSS model than in the chain ladder model, see Table 7.1. The variance of the predictive distribution is the mean squared error of prediction that we introduced in Chapter 1. We have a lot more uncertainty about the actual reserve estimate in the GAMLSS. Hence a business decision could be to add a margin on top of the reserve to reduce risk.
Against the background of Section 6.2 we believe that the GAMLSS estimate is more suitable than the chain ladder estimate in this case since the chain ladder method tends to underestimate claims even more than the GAMLSS. Especially for the last accident year the accident year reserve estimates are very different and affect the predictive distribution of the total reserve heavily. But since also the GAMLSS tends to underestimate cash flows one might think of using an even higher reserve in the end.

7.2. Line of Business 5

A detailed description of the GAMLSS for line of business 5 can be found in Appendix B.5. For reserve estimation, a Gaussian distribution has been used to model data (Section B.5.2). It is the only line of business where a common distribution could be used. Analyses of cash flow estimates showed that the chain ladder method tends to underestimate cash flows more than the GAMLSS model, compare Figure B.5.1. Hence we expect ultimate loss and total reserve estimates from the GAMLSS model to be higher than estimates from the chain ladder method.

Estimated ultimate losses from both methods behave very similarly (Figure 7.3). It is difficult to see a difference for the estimated means but the estimate for accident year 2011 from the GAMLSS simulation is slightly higher than the estimate from chain ladder bootstrapping. Estimates from the chain ladder bootstrapping vary more for older accident years, but for accident year 2011 the GAMLSS estimates spread wider around the estimated mean. Noticable is the wider spread around the mean for the GAMLSS at accident year 2003.
From these plots we don't expect the predictive reserve distributions to differ a lot. Also noticable is the observation that estimated ultimate losses from both lines of business decrease with increasing accident years. But the underlying data shows the same trend, i.e. a shrinking business for this line of business. Hence the observed trend is not caused by a lack of fit but in line with the underlying data.

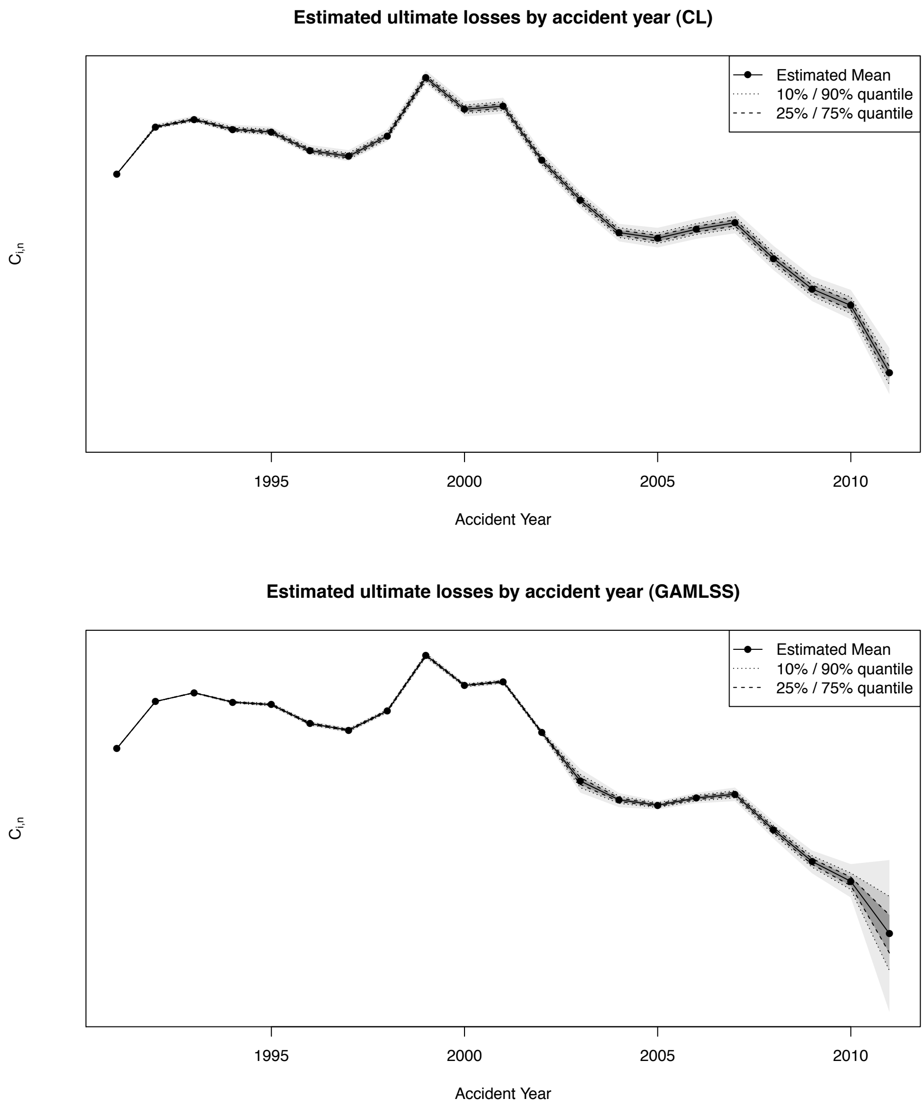


Figure 7.3.: Ultimate Loss estimates for both methods (LoB 5)

In Figure 7.4 predictive distributions of reserves are very similar between both methods. The empirical mean varys only slightly between both methods, which is caused by the different reserve estimate for accident year 2011.

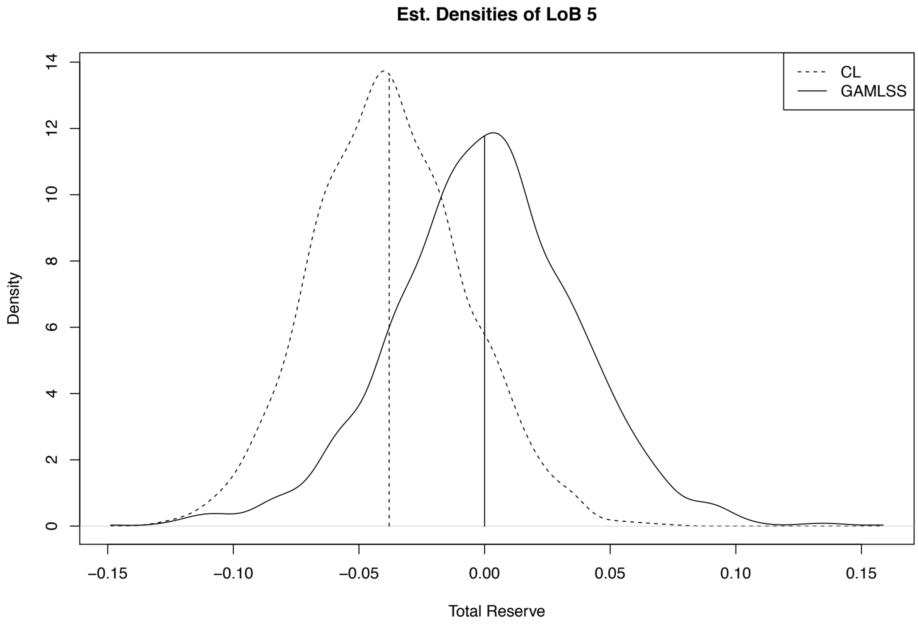


Figure 7.4.: Estimated densities for line of business 5

Statistics for both distribution are not too different, as the plot suggests. The empricial

	GAMLSS	CL
Mean	0.0000	-0.0380
Variance	0.0014	0.0009

Table 7.2.: LoB 5 total reserve statistics

mean from chain ladder bootstrapping is only 3.8% below the GAMLSS estimate. But again the empirical variance of the GAMLSS is greater than the one from chain ladder bootstrapping, although not as extreme as in for line of business 3. Since the chain ladder method tends to underestimate cash flows more than the GAMLSS, we prefer the GAMLSS model for this line of business. In contrast to line of business 3 the choice for one of the method almost mainly affects the amount of reserves which is put pack, because variances are not as different as for LoB 3.

7.3. Other Lines of Business

We summarize results for the other three lines of business briefly:

LoB1: Estimates of ultimate losses by accident year are similar for all accident years but 2011 (Figure B.1.7). For this year GAMLSS estimates are much higher and vary a lot more than chain ladder bootstrapping estimates. Thus the resulting predictive total reserve distribution from the GAMLSS model has a higher mean and much bigger variance than the one from chain ladder bootstrapping. Recalling results from Section 6.1, cash flows estimated by chain ladder method and GAMLSS where sometimes higher and sometimes lower than observed cash flows. Since cash flow estimation for calendar year 2011 shows significant underestimation, one could tend to choose the GAMLSS.

LoB2: Estimated ultimate losses are very similar for both methods for older accident years (Figure B.2.7). But the chain ladder bootstrapping estimates spread much wider around the estimated means and thus the predictive reserve distribution has a slightly higher variance than the one from the GAMLSS model. he estimated total reserve from chain ladder bootstrapping is only roughly 6.5% below the GAMLSS estimate. We also have to take observations from the cash flow analysis into account. The GAMLSS tends to overestimate losses while the chain ladder tends to underestimate them. Hence a value between both total reserve estimates makes sense.

LoB4: The GAMLSS estimates much higher reserves for all accident years, especially for the more recent accident years. Variation of estimates is not too different between both methods, but still the GAMLSS shows greater variation. Looking at the predictive distributions, the estimated mean from chain ladder bootstrapping is more than 30% below the one from GAMLSS, see Figure B.4.7. The cash flow analysis as shown in Appendix B.4 however gives no obvious explanation for this huge difference. Both methods tended to overestimate cash flows with none of them exceeding the other one for all calendar years. Here it is difficult to name the better model, so further research would have to be done here. On the other side, volume of this line of business compared to other lines of business in this portfolio is rather small. Hence an over- or underestimation of the reserve does not have a big influence on the total portfolio reserve.

7.4. Joint Reserve Estimation

So far estimation of reserves, either by accident year or total reserves per line of business has been done independently of results from other lines of business. That means total reserves for a portfolio can be estimated by summing up reserves of the different lines of business. This approach assumes that claims for each line of business are totally independent from other lines of business. While this is easy to model and estimate, it might not represent the given situation adequately. Maybe there is a dependence among different lines of business which affect the total reserve of a portfolio. This could either lead to higher portfolio reserves when there is a positive dependence or to lower portfolio reserves when there are diversification effects present. To examine dependencies among lines of business we follow the approach outlined in Section 5.3 and based upon Shi and Frees (2011). We have seen in Chapter 4 that for all lines of business GAMLSS with explicitly modeled scale parameter were necessary. In Section 5.3 this lead to a very small correlation between lines of business, which possibly could be neglected at all. There we also introduced two measures of dependence, Pearson's rho and Kendalls' tau. To follow up on the question why Pearson's rho is a good measure only in some situations, we give two examples where Pearson's rho fails. Kendall's tau does not have the same drawbacks since the dependence is measured on ranks of observations and not on observations and hence is more robust (e.g. against outliers).

Example (Non-Linear Dependence): Let $X \sim \mathcal{N}(0,1)$, i.e. let X be standard normal distributed. Then the first four moments of X are

$$\begin{aligned} \mathbb{E}[X] &= 0 \\ \mathbb{E}\left[X^2\right] &= 1 \\ \mathbb{E}\left[X^3\right] &= 0 \\ \mathbb{E}\left[X^4\right] &= 3. \end{aligned}$$

Let furthermore Y be a random variable defined as $Y := X^2$. Obviously Y strongly depends on X since it is a monotonic function of X. However, Pearson's rho for X and Y is

$$\rho(X,Y) = \rho(X,X^2) = \frac{\mathbb{E}[XX^2] - \mathbb{E}[X]\,\mathbb{E}[X^2]}{\sqrt{\sigma_X}\sqrt{\sigma_{X^2}}} = \frac{0-0}{\sqrt{1}\sqrt{2}} = 0.$$

Although a functional dependence is present, Pearson's rho is not able to capture it because it is a non-linear dependence between the two random variables. $\diamondsuit$

The second example uses the empirical versions of Pearson's rho and Kendall's tau. Let (x_i, y_i), $i = 1, \ldots, n$, be n pairs of observations from (X,Y). Then Pearson's rho (the sample correlation) is calculated as

$$\hat{\rho}(X,Y) = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}},$$

where $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$ are the sample means of X and Y, respectively.
The empirical version for Kendall's tau uses the number of concordant and discordant

pairs. Let (x_i, y_i), $i = 1, \ldots, n$, be n pairs of observations from (X, Y). For $i \neq j$ two pairs (x_i, y_i) and (x_j, y_j) are said to be

(i) **concordant** if

$$\big(x_i > x_j \text{ and } y_i > y_j\big) \quad \text{or} \quad \big(x_i < x_j \text{ and } y_i < y_j\big)$$

(ii) **discordant** if

$$\big(x_i > x_j \text{ and } y_i < y_j\big) \quad \text{or} \quad \big(x_i < x_j \text{ and } y_i > y_j\big)$$

(iii) **tied** if

$$\big(x_i \neq x_j \text{ and } y_i = y_j\big) \quad \text{or} \quad \big(x_i = x_j \text{ and } y_i \neq y_j\big) \quad \text{or} \quad \big(x_i = x_j \text{ and } y_i = y_j\big).$$

Let

$$\begin{aligned} N_{con} &= \#\text{ concordant pairs} \\ N_{dis} &= \#\text{ discordant pairs} \\ N_{xt} &= \#\text{ tied pairs with tie in the x's} \\ N_{yt} &= \#\text{ tied pairs with tie in the y's.} \end{aligned}$$

Then Kendall's tau is calculated by

$$\hat{\tau}(X, Y) = \frac{N_{con} - N_{dis}}{\sqrt{N_{con} + N_{dis} + N_{xt}}\sqrt{N_{con} + N_{dis} + N_{yt}}}.$$

Example (Outlier): Let x and y be two vectors of length 1000. Let the first 999 entries of both vectors be random numbers from a standard normal distribution and let the 1000th entry be 10000.

```
> x <- c(rnorm(999), 10000)
> y <- c(rnorm(999), 10000)
```

Then Pearson's rho is

```
> cor(x,y, method='pearson')
```

```
[1] 0.9999897
```

while Kendall's tau is

```
> cor(x,y, method='kendall')
```

```
[1] -0.0162002
```

The outlier distorts Pearson's rho heavily. Here Kendall's tau is much more robust than Pearson's rho since it is a rank correlation, i.e. only the number of concordant and discordant pairs are used for the calculation but not the x- and y-values themselves. $\diamondsuit$

These two examples showed that Pearson's rho is not a good dependence measure in every situation. Since even some more problems can arise when using Pearson's rho we use Kendall's tau to examine dependencies among lines of business.

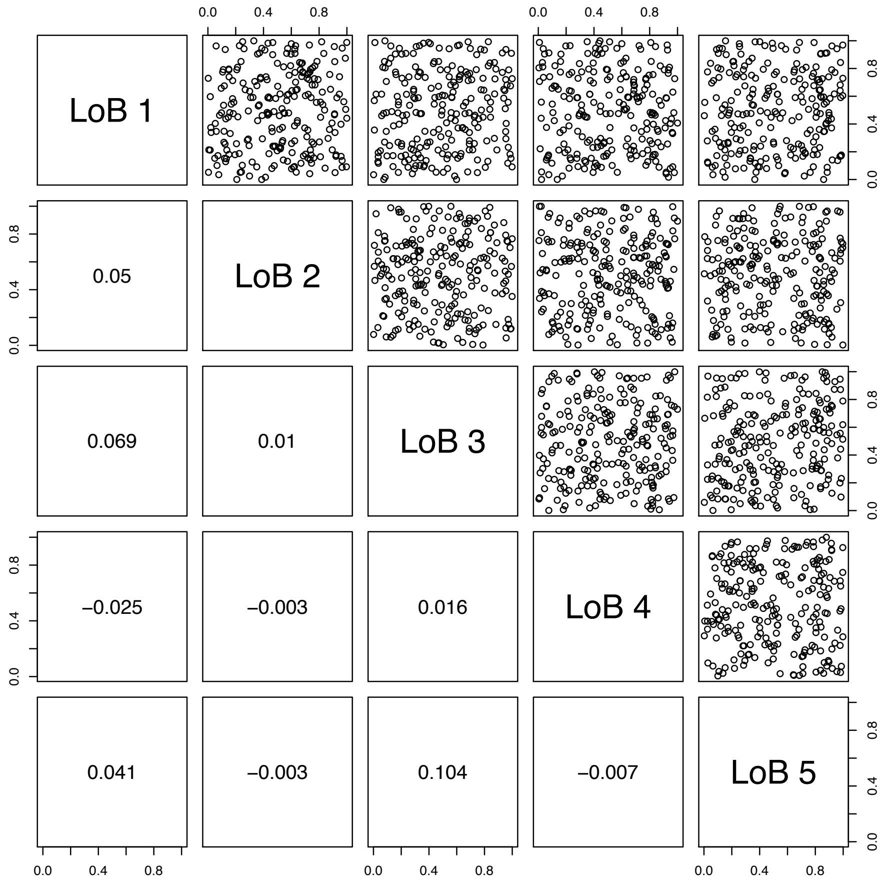


Figure 7.5.: Scatterplot and Kendall's tau's of residuals

Figure 7.5 shows a scatterplot of residuals from the GAMLSS in the upper triangle. The lower triangle shows Kendall's tau for each pair of lines business. Surprisingly all values are very close to zero, suggesting that residuals of the GAMLSS models are almost independent between lines of business. Only between line of business 3 and 5 there seems to be a dependence with Kendall's tau being 0.104. For all other pairs of lines of business correlation is rather small. Note that four negative estimates occur and diversification effects could appear when modeling joint reserve estimates.

Figure 7.5 suggests that dependencies among different lines of business in this portfolio don't seem to be present when looking at residuals of the GAMLSS models. But concluding that the lines of business are completely independent is too premature. For each model between 21 and 24 parameters had been used, 9 - 11 solely for modeling the scale parameter. By that almost every systematic pattern in data has been captured by the models. Hence residuals should not include any pattern and correlation between residuals of different lines of business is expected to be very small or not present at all.

7.5. Portfolio Reserve

As outlined in the last section observed Kendalls's tau's are very small. Almost no dependence is present in this setup and thus there is no need to fit a copula model to incorporate dependencies. The use of a copula model would not affect the joint ultimate losses and reserves a lot.

We therefore follow the traditional way of summing up reserves from all lines of business in the portfolio to gain the portfolio reserve. While this is similar to what one would do with chain ladder estimates of reserves, the preconditions are very different. The chain ladder method simply assumes independence of different lines of business in a portfolio and no checks to verify this are done. In contrast to that, we incorporated dependencies and examined them on residuals of the GAMLSS. Not till we showed that dependencies among residuals are very small we used the assumption of having independent lines of business in the portfolio.

Figure 7.6 shows ultimate losses of the portfolio by accident year. Except for the peak in accident year 2002 the curves look very similar to the ones from line of business 5 (see Figure 7.3). This is caused by the fact that line of business 5 has much more volume than all other lines of business, i.e. more premiums and claims. Hence ultimate losses of this small portfolio are dominated by line of business 5. As shown in Appendix B.1.2 the peak for accident year 2002 is caused by line of business 1.

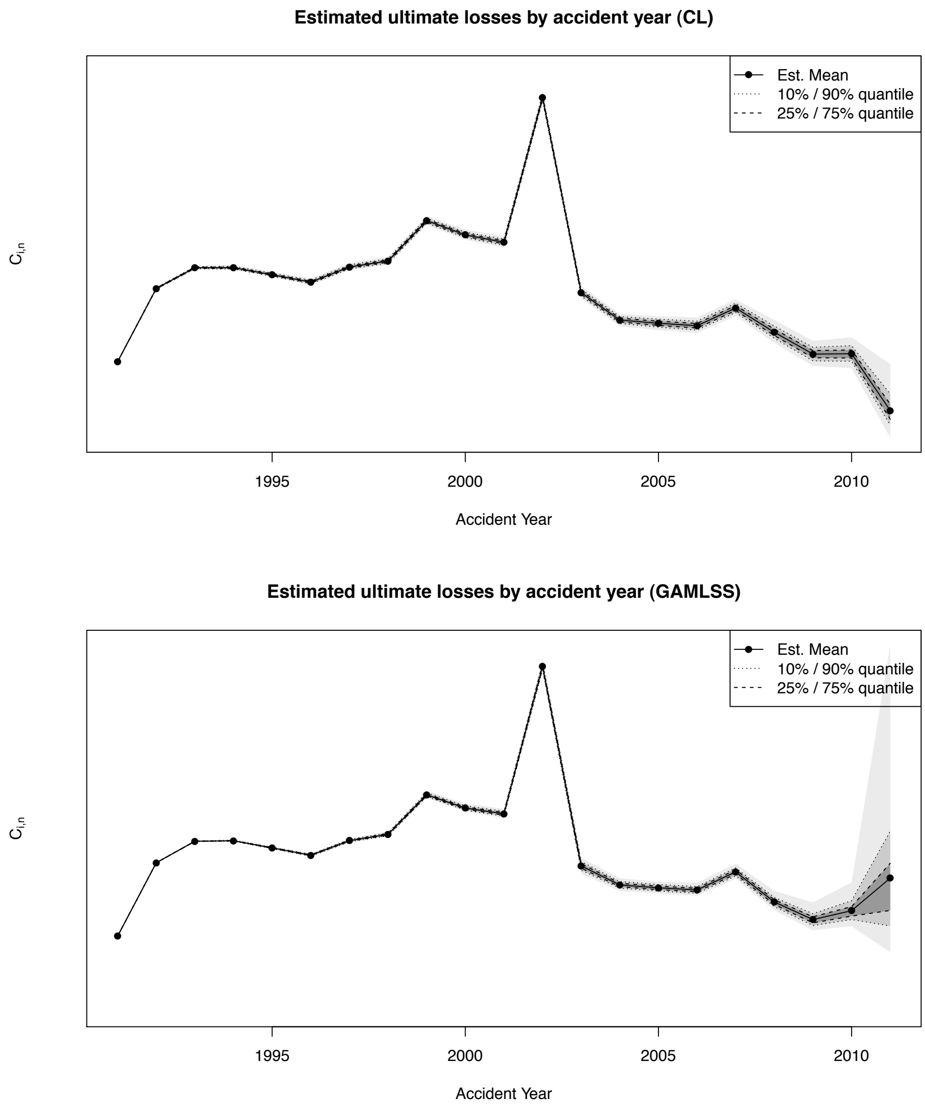


Figure 7.6.: Ultimate loss estimates for both methods of the portfolio

Since line of business 5 has the biggest influence on ultimate claims, we expect the density plot of total reserves to look similar to Figure 7.4. Figure 7.7 shows at least some similarities for the shape of the densities. But the empirical variance differs a lot more than for LoB 5.

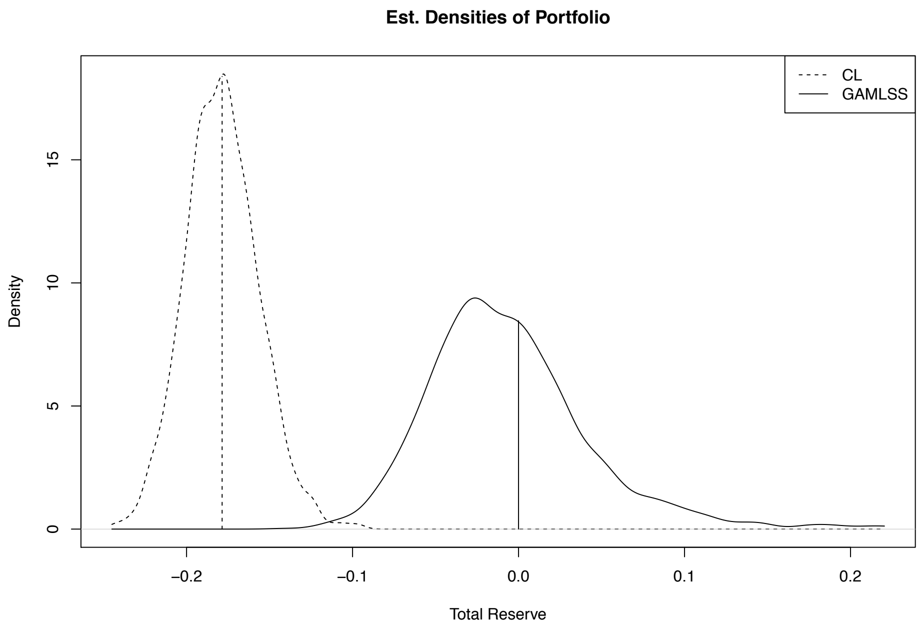


Figure 7.7.: Estimated densities for portfolio reserve

	GAMLSS	CL
Mean	0.0000	-0.1786
Variance	0.0060	0.0005

Table 7.3.: Portfolio total reserve statistics

The empirical mean of estimated reserves from chain ladder bootstrapping is 17.86% below the empirical mean of estimated reserves from the GAMLSS simulations. The variance is much bigger for the GAMLSS (about 12 times). This is mainly caused by the huge standard deviation from line of business 1.
Hence there is a big difference between both methods. The estimate from chain ladder bootstrapping is 17.86% below the GAMLSS estimate. Although this is a big difference in absolut values, it is basically that is just a number the insurance company would have to write in their book. The more interesting result is, that the variance is much greater and hence uncertainty and risk is much bigger than the chain ladder bootstrapping suggests. This could lead to problems for the company if the ultimate loss is higher than expected. Hence it would have been even more eligible to have diversification effects in the portfolio and reduce variation. Anyway, the insurance company would need to think of actions to take to be able pay higher losses than expected. They are much more probable when using a GAMLSS rather than the chain ladde method and the company would have to be prepared for that.

8. Conclusion and Outlook

In this thesis we introduced generalized additive models for location, scale and shape as a statistical model for loss triangles. They allow for explicit modeling of the scale parameter, which is often closely related to the variance. By that a better fit than with generalized linear models can be achieved and it is necessary to rather use a GAMLSS than GLM in many situations.
Inference about ultimate losses and reserves is possible in this setup and can be compared to established methods like the chain ladder method. As for generalized linear models, predictive distributions can be obtained and uncertainty can be measured. Complex distributions like zero-inflated distributions or the SEP1 distribution allow for a maximum of flexibility and guarantee that each line of business can be modeled in the best possible way. Estimates of next calendar year's cash flow were closer to the observed cash flows for the GAMLSS in many cases. But no general trend was observable and it depends on the underlying data set which method should be used. Predictive ultimate loss distributions from the GAMLSS had a greater variance in most cases. Hence uncertainty is greater when using a GAMLSS on the one hand side. On the other side this can be a warning, such that ultimate losses can be much higher than current methods like the chain ladder method suggest.
We examined dependencies of lines of business in an insurance portfolio among residuals of the GAMLSS. Since Kendall's tau was rather small, we argue that not much dependence of lines of business is present in the portfolio. Therefore we conclude that dependencies in an insurance portfolio are secondary if each lines of business is modeled properly by a GAMLSS. This implicates that portfolio reserves can be estimated in a similar way as for the chain ladder method by summing up reserves of individual lines of business.

For the dependence analysis we have not considered all available lines of business. For other lines of business zero-inflated distributions were necessary and a different approach to examine dependencies would be required. Then dependencies in a bigger portfolio could be analyzed and different outcomes are possible. Hence the approaches and findings presented in this thesis can serve as a starting point to analyze loss triangles and dependencies within a big insurance portfolio. The GAMLSS plays a key role as a more sophisticated statistical model than the generalized linear model and thus can improve analyses, results and business decisions significantly.

A. Inflation and Dependence Models

A.1. Inflation

This section gives details about the two GAMLSS used in Section 5.2.

The triangle with incremental payments used in this section has a claims history of 26 years. Inflation effects are present and the triangle has been manually deflated using a consumer price index. The chain ladder method has been applied to both triangles in the known way to estimate the cash flow. Here, the cash flow is defined as

$$S_{n+1} = S_{2,n} + S_{3,n-1} + \ldots + S_{n,2}.$$

$S_{1,n+1}$ is not part of the cash flow since no payment after the last development is expected.

The GAMLSS models Paid-to-Premium ratios which don't need to be deflated. Inflation effects cancel down automatically and hence only one GAMLSS needs to be fitted:

$$\begin{aligned} PtP.Ratio_{i,k} \sim \mathrm{SEP1}\big(\mu_{i,k} &= poly(DevLag_k, 3) + poly(AYear_i, 3) \\ &\quad + (DevLag_k = 1) + (DevLag_k = 2) \\ &\quad + (DevLag_k = 3) + (DevLag_k = 4) \\ &\quad + (DevLag_k = 7) + (DevLag_k = 8) \\ \log(\sigma_{i,k}) &= poly(DevLag_k, 4) + AYear_i \\ &\quad + (DevLag_k = 3) + (DevLag_k = 4), \\ &\quad + (DevLag_k = 7) + (DevLag_k = 8), \\ &\quad + (AYear_i = 1986), \\ \nu_{i,k} &= 1, \\ \log(\tau_{i,k}) &= 1\big). \end{aligned}$$

The model uses a skew exponential power type 1 distribution and has 25 parameters. Residual plots shown in Figure A.1.1 and show no lack of fit.

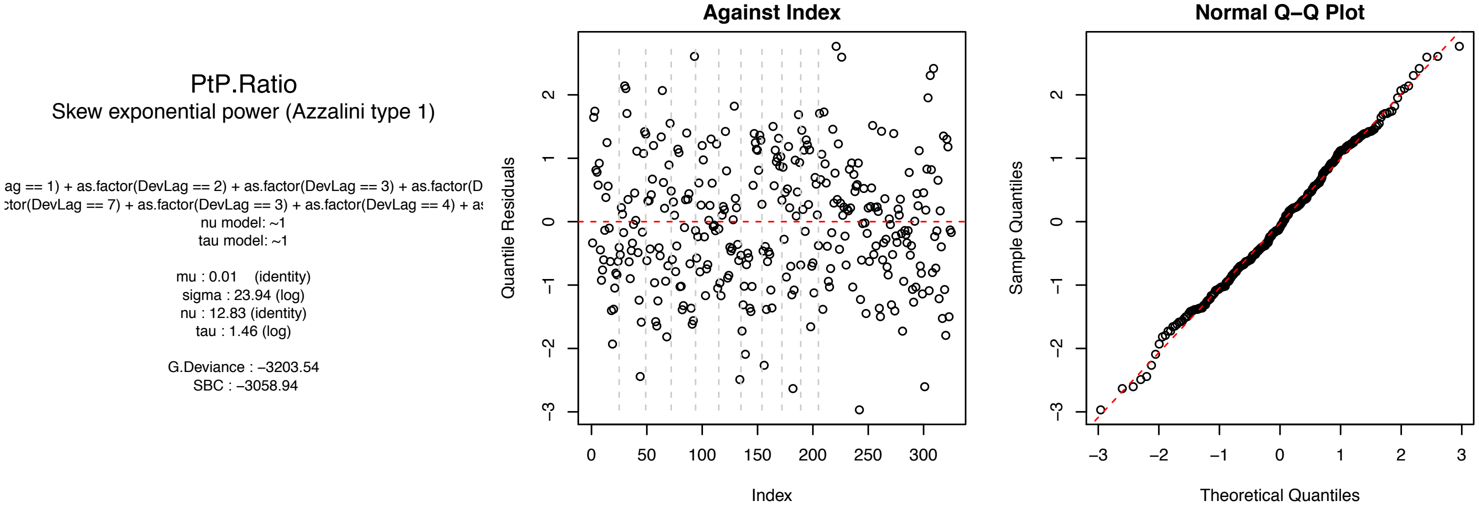


Figure A.1.1.: Residuals Plots for GAMLSS applied to Paid-to-Premium ratios

A.2. Dependence

In this section we give some details about the two GLMs and GAMLSS used in Section 5.3.

Triangles are given as in Table A.2.1 and Table A.2.2. Both triangles include 10 years of history and thus 55 observations. Premiums are used to calculate the Paid-to-Premium ratios and fit GLMs and GAMLSS to the PtP ratios.

Accident Year i	Premium	Development Lag k									
		1	2	3	4	5	6	7	8	9	10
1	4,711,333	1,376,384	1,211,168	535,883	313,790	168,142	79,972	39,235	15,030	10,865	4,086
2	5,335,525	1,576,278	1,437,150	652,445	342,694	188,799	76,956	35,042	17,089	12,507	
3	5,947,504	1,763,277	1,540,231	678,959	364,199	177,108	78,169	47,391	25,288		
4	6,354,197	1,779,698	1,498,531	661,401	321,434	162,578	84,581	53,449			
5	6,738,172	1,843,224	1,573,604	613,095	299,473	176,842	106,296				
6	7,079,444	1,962,385	1,520,298	581,932	347,434	238,375					
7	7,254,832	2,033,371	1,430,541	633,500	432,257						
8	7,739,379	2,072,061	1,458,541	727,098							
9	8,154,065	2,210,754	1,517,501								
10	8,435,918	2,206,886									

Table A.2.1.: Commercial Auto (incremental paid losses)

Accident Year i	Premium	Development Lag k									
		1	2	3	4	5	6	7	8	9	10
1	267666	33,810	45,318	46,549	35,206	23,360	12,502	6,602	3,373	2,373	778
2	274526	37,663	51,771	40,998	29,496	12,669	11,204	5,785	4,220	1,910	
3	268161	40,630	56,318	56,182	32,473	15,828	8,409	7,120	1,125		
4	276821	40,475	49,697	39,313	24,044	13,156	12,595	2,908			
5	270214	37,127	50,983	34,154	25,455	19,421	5,728				
6	280568	41,125	53,302	40,289	39,912	6,650					
7	344915	57,515	67,881	86,734	18,109						
8	371139	61,553	132,208	20,923							
9	323753	112,103	33,250								
10	221448	37,554									

Table A.2.2.: Personal Auto (incremental paid losses)

The summaries for both models are shown below. For Personal Auto parameter estimates are the same as in Shi and Frees (2011), while for Commercial Auto small differences can be seen.

Personal Auto:

```
The following object(s) are masked _by_ '.GlobalEnv':

    AYear
******************************************************************
Family:  c("LOGNO", "Log Normal")

Call:
gamlss(formula = Personal.Auto ~ as.factor(DevLag) + as.factor(AYear),
    sigma.formula = ~1, family = LOGNO(mu.link = "identity",
        sigma.link = "identity"), data = shifrees, method = RS(1000),
    control = gamlss.control(mu.step = 0.01, sigma.step = 0.01,
        gd.tol = 100))

Fitting method: RS(1000)

------------------------------------------------------------------
Mu link function:  identity
Mu Coefficients:
                      Estimate  Std. Error   t value   Pr(>|t|)
(Intercept)           -1.13674     0.04284  -26.5341  3.145e-25
as.factor(DevLag)2    -0.22443     0.04182   -5.3672  4.873e-06
as.factor(DevLag)3    -1.04691     0.04373  -23.9393  1.053e-23
as.factor(DevLag)4    -1.64405     0.04582  -35.8779  8.923e-30
as.factor(DevLag)5    -2.25397     0.04831  -46.6550  8.492e-34
as.factor(DevLag)6    -3.01297     0.05147  -58.5432  2.664e-37
as.factor(DevLag)7    -3.67129     0.05575  -65.8565  3.997e-39
as.factor(DevLag)8    -4.49346     0.06211  -72.3521  1.386e-40
as.factor(DevLag)9    -4.91091     0.07302  -67.2511  1.891e-39
as.factor(DevLag)10   -5.91342     0.09851  -60.0300  1.090e-37
as.factor(AYear)1989  -0.03273     0.04182   -0.7828  4.389e-01
as.factor(AYear)1990  -0.02844     0.04373   -0.6503  5.196e-01
as.factor(AYear)1991  -0.13087     0.04582   -2.8559  7.082e-03
as.factor(AYear)1992  -0.17467     0.04831   -3.6156  9.100e-04
as.factor(AYear)1993  -0.17446     0.05147   -3.3899  1.709e-03
as.factor(AYear)1994  -0.17295     0.05575   -3.1023  3.723e-03
as.factor(AYear)1995  -0.22337     0.06211   -3.5966  9.601e-04
as.factor(AYear)1996  -0.24436     0.07302   -3.3463  1.927e-03
as.factor(AYear)1997  -0.20417     0.09851   -2.0727  4.542e-02

------------------------------------------------------------------
Sigma link function:  identity
```

```
Sigma Coefficients:
             Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)   0.08865    0.008458    10.48  1.259e-14

-------------------------------------------------------------------
No. of observations in the fit:  55
Degrees of Freedom for the fit:  20
      Residual Deg. of Freedom:  35
                      at cycle:  6

Global Deviance:     -435.1066
            AIC:     -395.1066
            SBC:     -354.9599
*******************************************************************
```

Commercial Auto:

```
The following object(s) are masked _by_ '.GlobalEnv':

    AYear
*******************************************************************
Family:  c("GA", "Gamma")

Call:
gamlss(formula = Commercial.Auto ~ as.factor(DevLag) + as.factor(AYear),
    sigma.formula = ~1, family = GA("inverse", sigma.link = "identity"),
    data = shifrees, method = RS(100), control = gamlss.control(mu.step = 0.01,
        sigma.step = 0.01, gd.tol = 100))

Fitting method: RS(100)

-------------------------------------------------------------------
Mu link function:  inverse
Mu Coefficients:
                       Estimate  Std. Error   t value   Pr(>|t|)
(Intercept)             5.80444      1.1037   5.25892  6.796e-06
as.factor(DevLag)2     -0.84210      0.8284  -1.01652  3.162e-01
as.factor(DevLag)3      0.32628      0.9759   0.33434  7.401e-01
as.factor(DevLag)4      3.33201      1.3383   2.48970  1.754e-02
as.factor(DevLag)5     11.58664      2.4515   4.72629  3.457e-05
as.factor(DevLag)6     20.65723      3.9465   5.23436  7.329e-06
as.factor(DevLag)7     42.16133      7.8151   5.39485  4.476e-06
as.factor(DevLag)8     87.34475     17.3573   5.03216  1.362e-05
as.factor(DevLag)9    120.26302     28.7967   4.17628  1.797e-04
as.factor(DevLag)10   338.23097    110.2571   3.06766  4.082e-03
as.factor(AYear)1989    0.66382      1.4242   0.46609  6.440e-01
as.factor(AYear)1990   -0.33100      1.3107  -0.25254  8.021e-01
```

```
as.factor(AYear)1991    1.08012      1.4747   0.73244  4.686e-01
as.factor(AYear)1992    1.05580      1.4742   0.71618  4.785e-01
as.factor(AYear)1993    0.56420      1.4220   0.39678  6.939e-01
as.factor(AYear)1994   -0.18798      1.3446  -0.13981  8.896e-01
as.factor(AYear)1995   -0.39760      1.3624  -0.29185  7.721e-01
as.factor(AYear)1996   -0.88923      1.4178  -0.62718  5.345e-01
as.factor(AYear)1997    0.09234      2.1965   0.04204  9.667e-01

-------------------------------------------------------------------
Sigma link function:  identity
Sigma Coefficients:
             Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)     0.322     0.03019    10.67  6.609e-15

-------------------------------------------------------------------
No. of observations in the fit:  55
Degrees of Freedom for the fit:  20
      Residual Deg. of Freedom:  35
                      at cycle:  6

Global Deviance:     -255.497
            AIC:     -215.497
            SBC:     -175.3504
*******************************************************************
```

Residuals plots for the GLMs and GAMLSS are shown in Figure A.2.1 and Figure A.2.2.

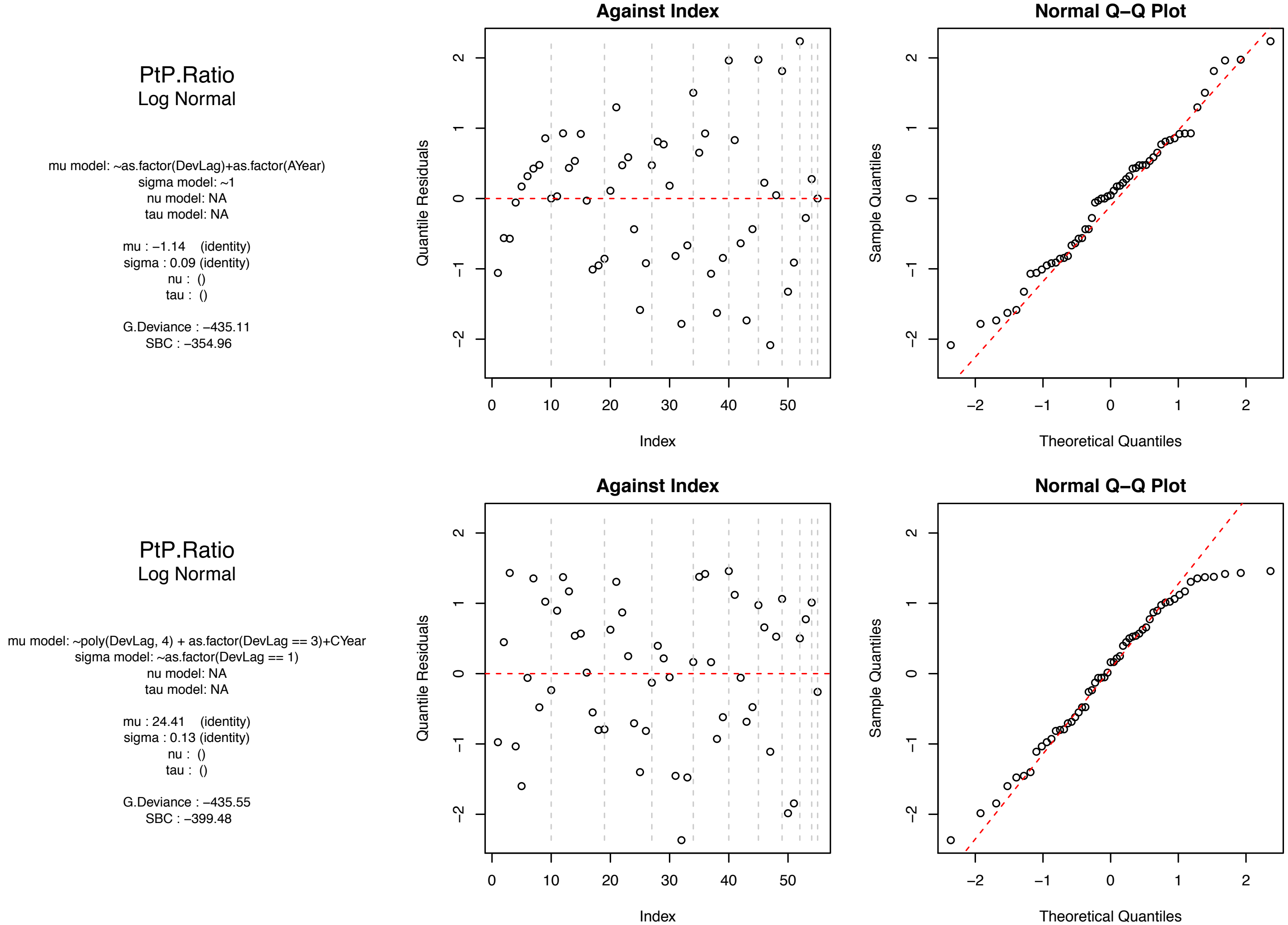


Figure A.2.1.: GLM and GAMLSS for Personal Auto

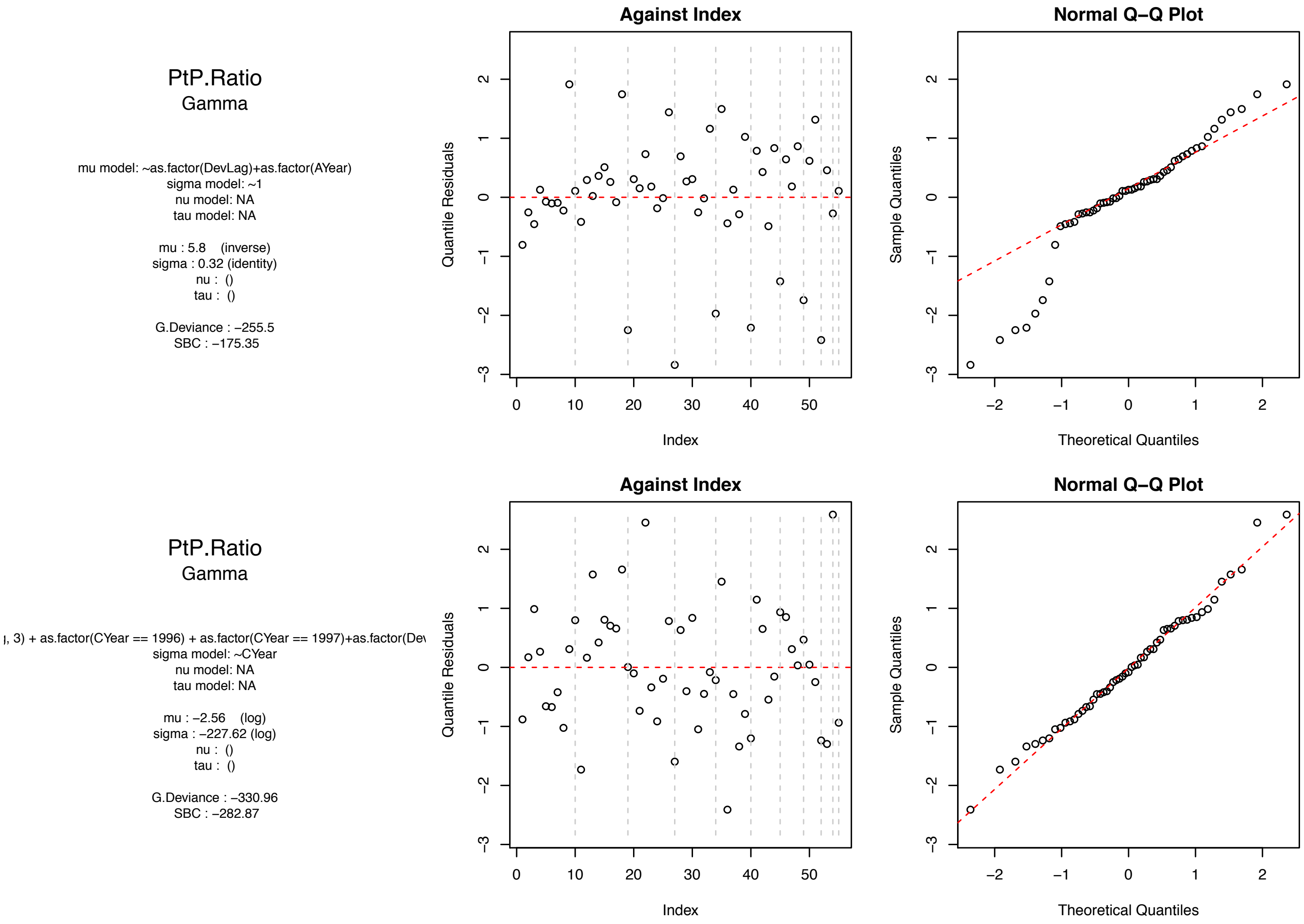


Figure A.2.2.: GLM and GAMLSS for Commercial Auto

B. Exhibits per Line of Business

B.1. Line of Business 1

Line of business 1 is a mid-sized line of business. Data is available for 21 years and has just a few peculiarities. Paid claims for accident year 2002 are much higher than for all other accident years, but development is similar to other accident years. There are only some negative payments but no zero-payments.

B.1.1. Cash Flow Analysis

Cash flow analysis shows very mixed results for this line of business, see Figure B.1.1. A detailed discussion of results can be found in Section 6.1.

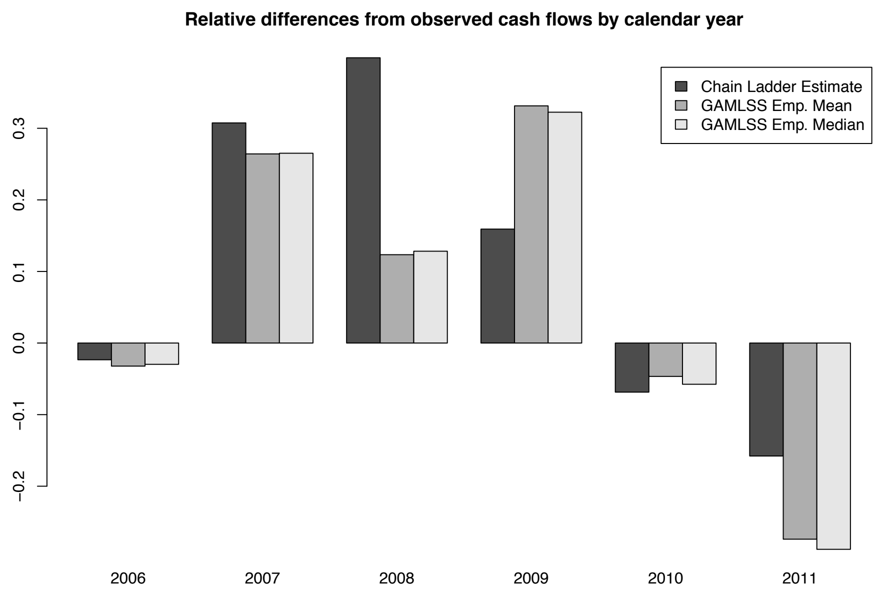


Figure B.1.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 1

Model	Distr.	Parameter	Polynomials	Factors	Interactions
2005	SEP1	μ	poly(DevLag,2)	DevLag 1,2,3	
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2)	AYear 2002, DevLag 3	
		ν	1		
		$\ln(\tau)$	1		
2006	SEP1	μ	poly(DevLag,2)	DevLag 1,2,3	(DevLag 3):poly(AYear,2)
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2)	AYear 2002, DevLag 3,7	
		ν	1		
		$\ln(\tau)$	1		
2007	SEP1	μ	poly(DevLag,2)	DevLag 1,2,3,4,5	(DevLag 3):poly(AYear,3)
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2), $(\text{CYear})^2$, poly(CYear,2)	AYear 2002, DevLag 3	
		ν	1		
		$\ln(\tau)$	1		
2008	SEP1	μ		DevLag 1,2,3,4,5	(DevLag 3):AYear
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2), $(\text{CYear})^2$	AYear 2002, DevLag 6,12	
		ν	1		
		$\ln(\tau)$	1		
2009	SEP1	μ	poly(DevLag,2)	DevLag 1,2,3,4,5	(DevLag 3):poly(AYear,3)
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2), $(\text{CYear})^2$	AYear 2002, DevLag 3, CYear 2008	
		ν	1		
		$\ln(\tau)$	1		
2010	SEP1	μ		DevLag 1,2,3,4,5	(DevLag 3):poly(AYear, 2), (DevLag 2):poly(AYear, 2)
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,2)	AYear 1997,2002, DevLag 3, CYear 2008	
		ν	1		
		$\ln(\tau)$	1		

Table B.1.1.: All six models for LoB 1

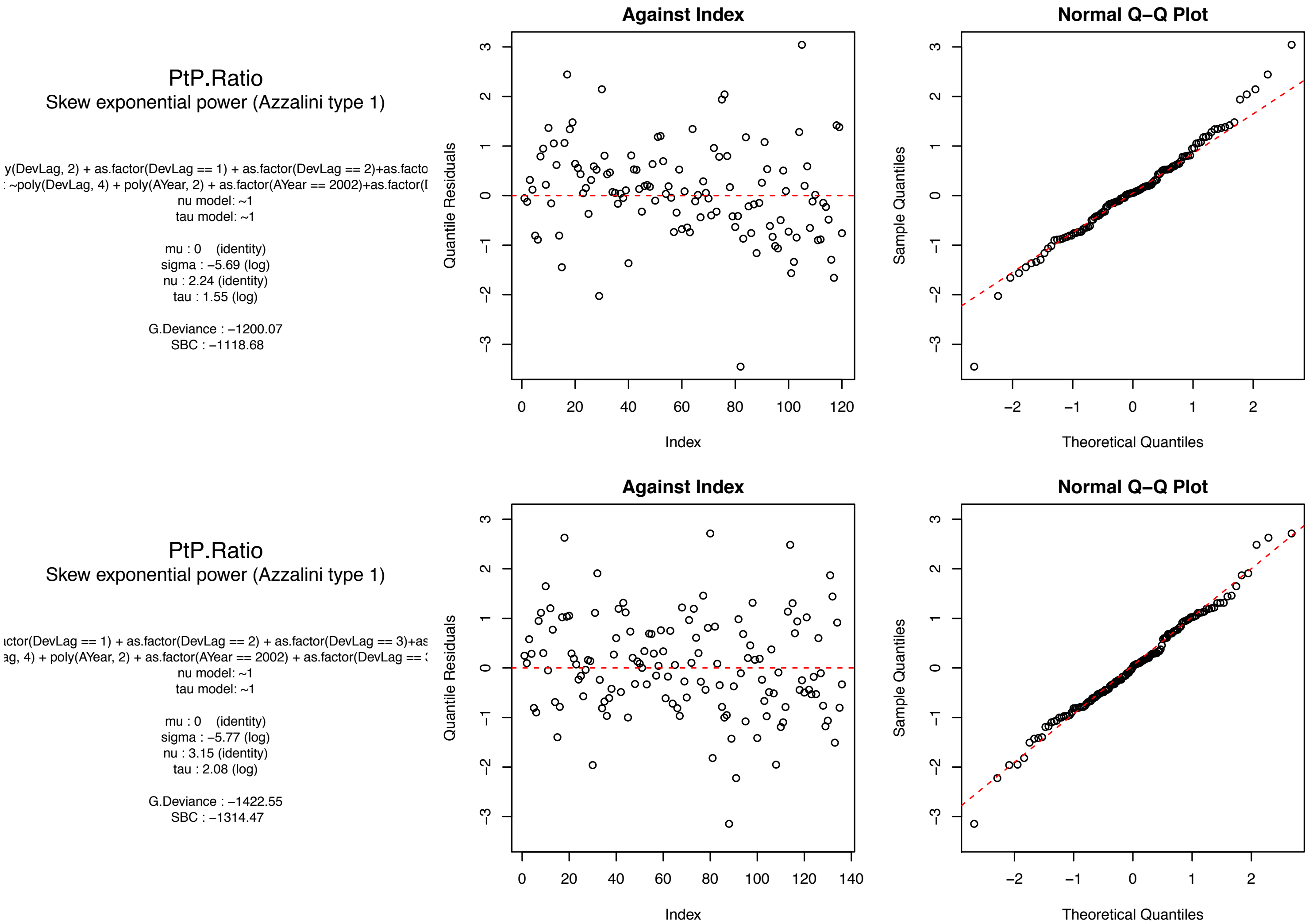


Figure B.1.2.: Residuals plots for LoB 1 (CYear 2005 & 2006)

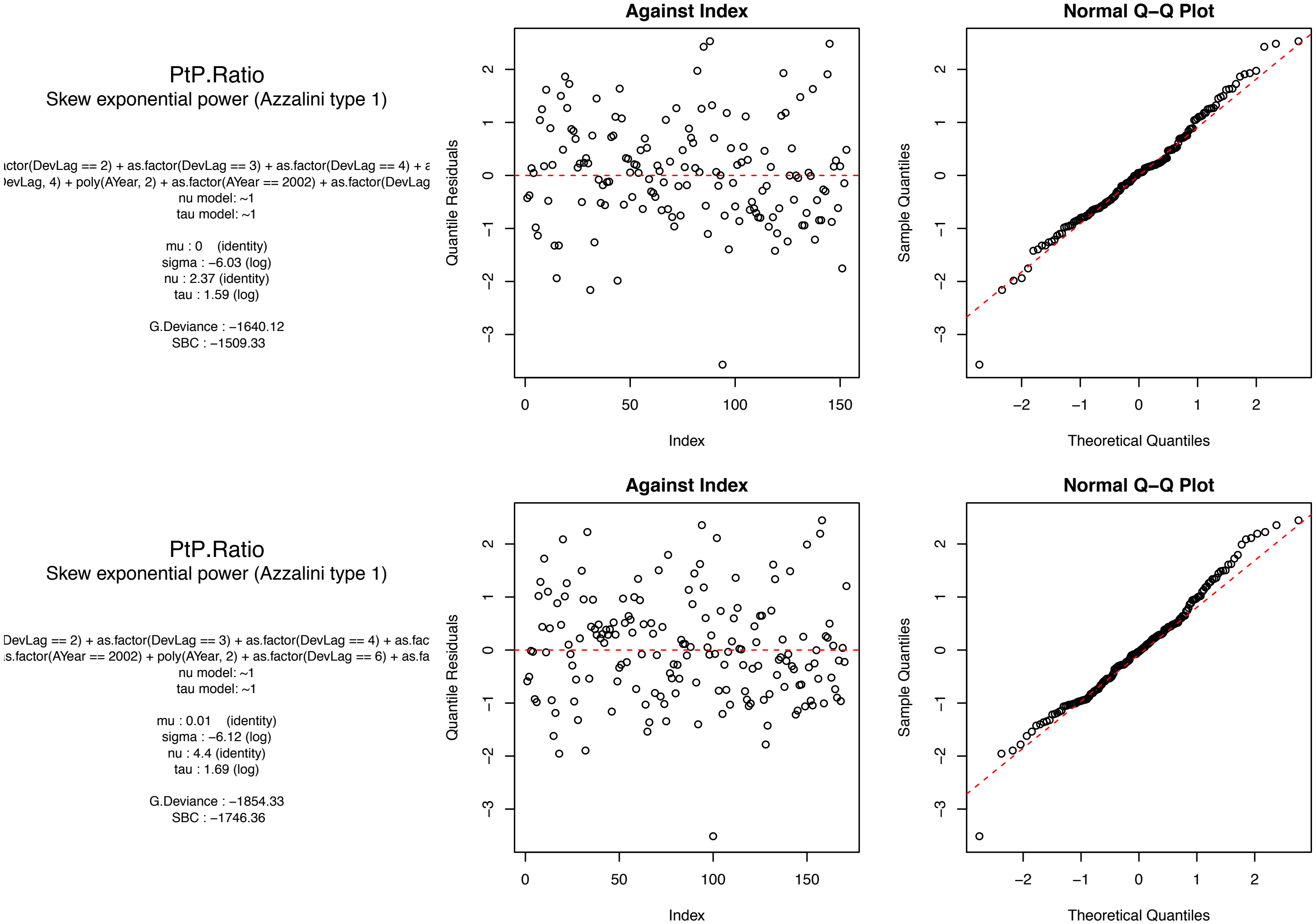


Figure B.1.3.: Residuals plots for LoB 1 (CYear 2007 & 2008)

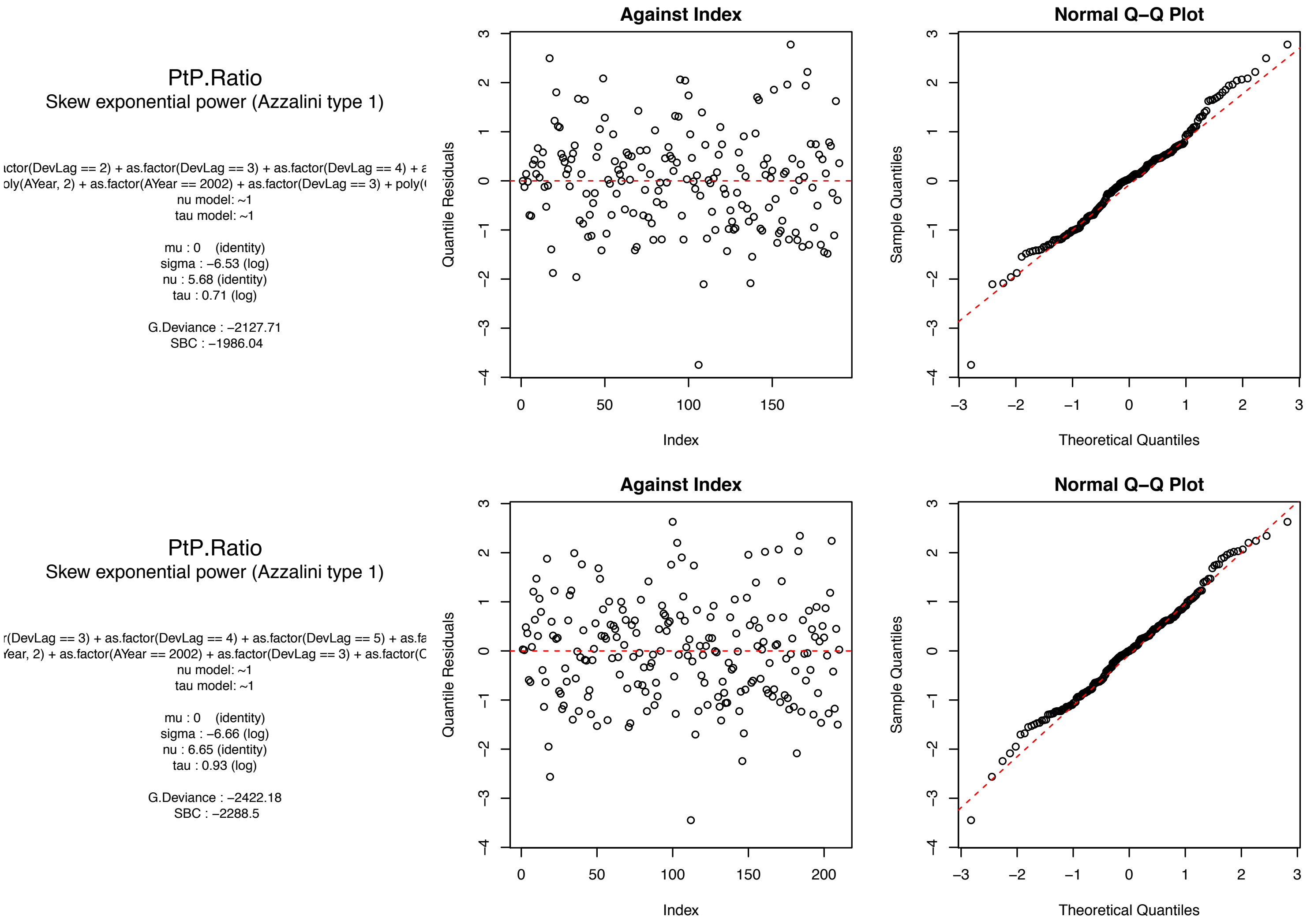


Figure B.1.4.: Residuals plots for LoB 1 (CYear 2009 & 2010)

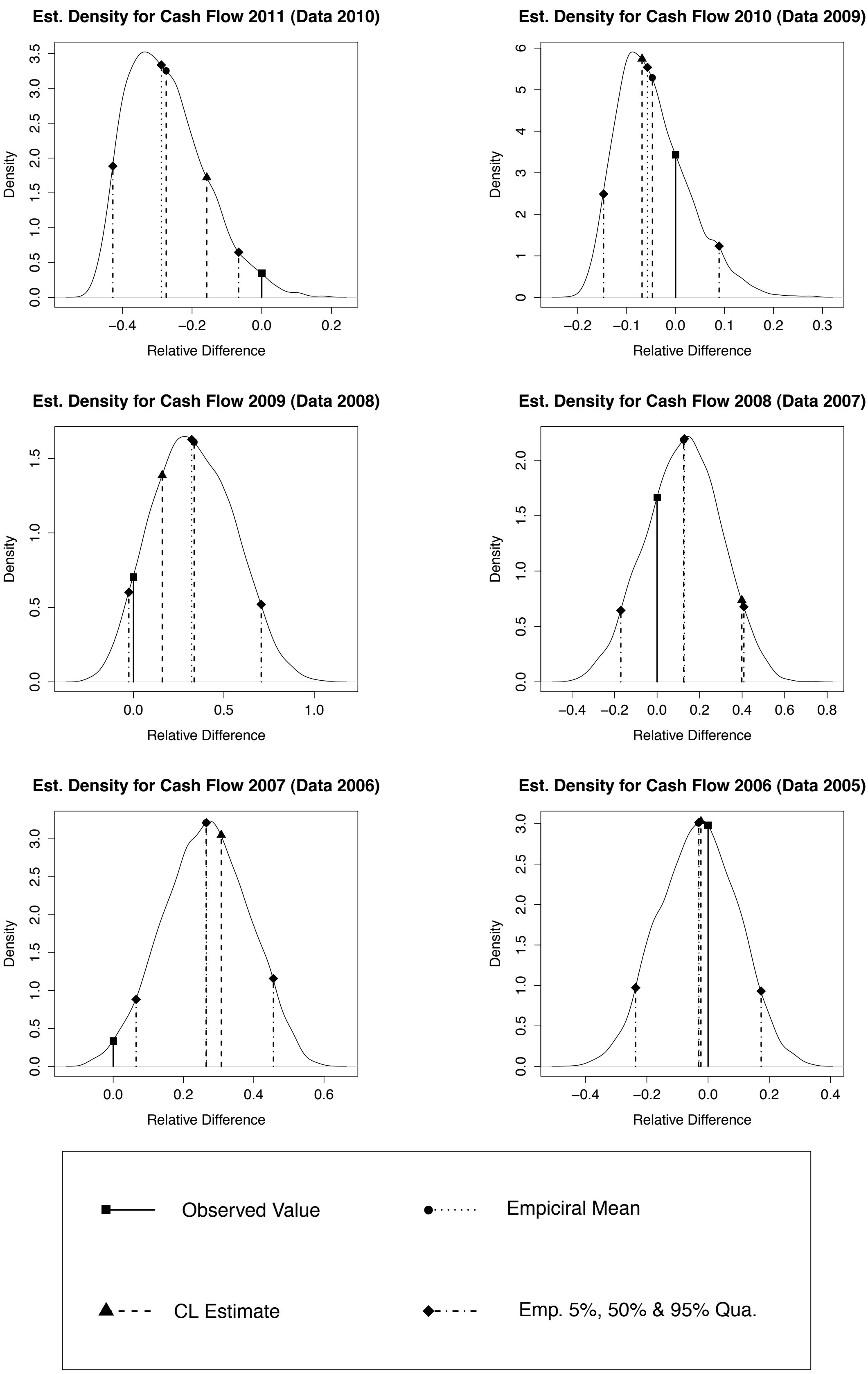


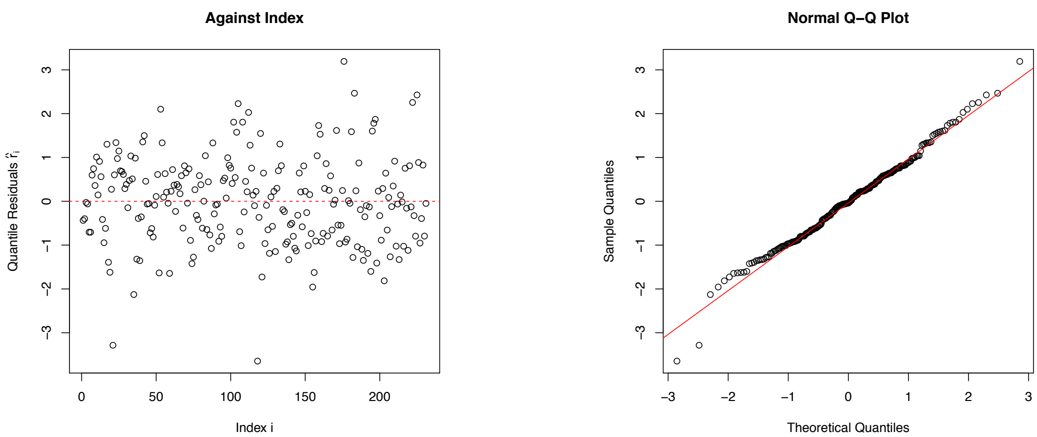
Figure B.1.5.: Predictive distributions for LoB 1

B.1.2. Latest Model & Reserving

The model using most recent data is

$$
\begin{aligned}
PtP.Ratio_{i,k} \sim \text{SEP1}\Big(\quad \mu &= DevLag_k + (DevLag_k = 1 + (DevLag_k = 2) \\
&\quad + (DevLag_k = 3) + (DevLag_k = 3) : poly(AYear_i, 2), \\
\ln(\sigma) &= poly(DevLag_k, 4) + poly(AYear_i, 2) + poly(CYear_t, 2) \\
&\quad + (AYear_i = 2002) + (DevLag_k = 3) + (CYear_t = 2008), \\
\nu &= 1, \\
\ln(\tau) &= 1 \Big).
\end{aligned}
$$

Residuals show no lack of fit:



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure B.1.6.: Residual plots for Line of Business 1

Statistics of reserve distributions for lines of business 1:

	GAMLSS	CL
Mean	0.0000	-0.8425
Variance	0.5433	0.0003

Table B.1.2.: LoB 1 total reserve statistics

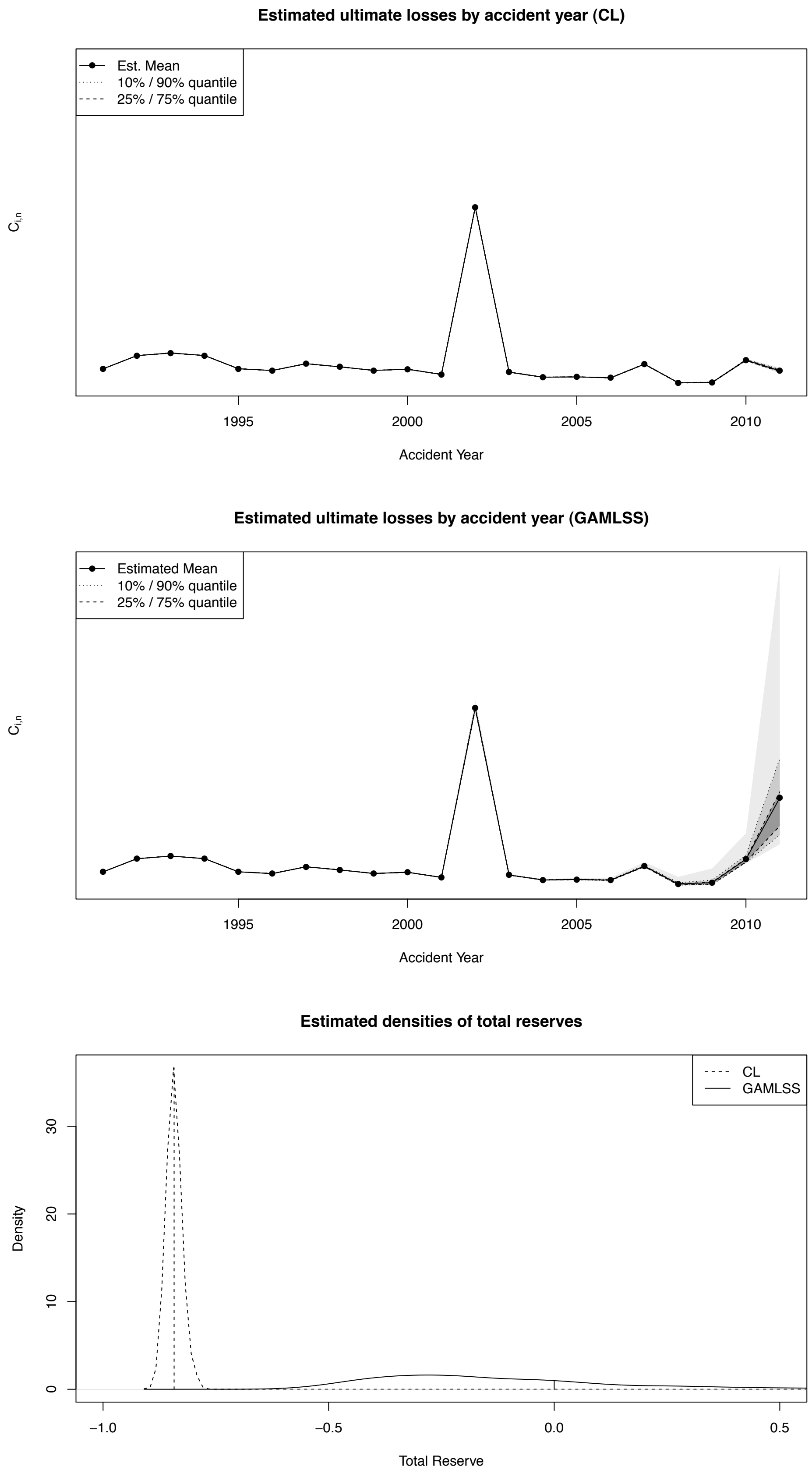


Figure B.1.7.: Ultimate loss estimates and estimated total reserve densities (LoB 1)

B.2. Line of Business 2

B.2.1. Cash Flow Analysis

Cash flow analysis (see Figure B.2.1) shows that while the GAMLSS tends to overestimate the claims, chain ladder tends to underestimate. Absolute values of chain ladder estimates are smaller than absolute values of GAMLSS estimates for four years. Hence one could argue that the chain ladder method is more appropriate than the GAMLSS. We argue that underestimation of losses is more risky for an insurer than overestimation. Thus we think the GAMLSS provides more conservative and 'better' estimates of the cash flows.

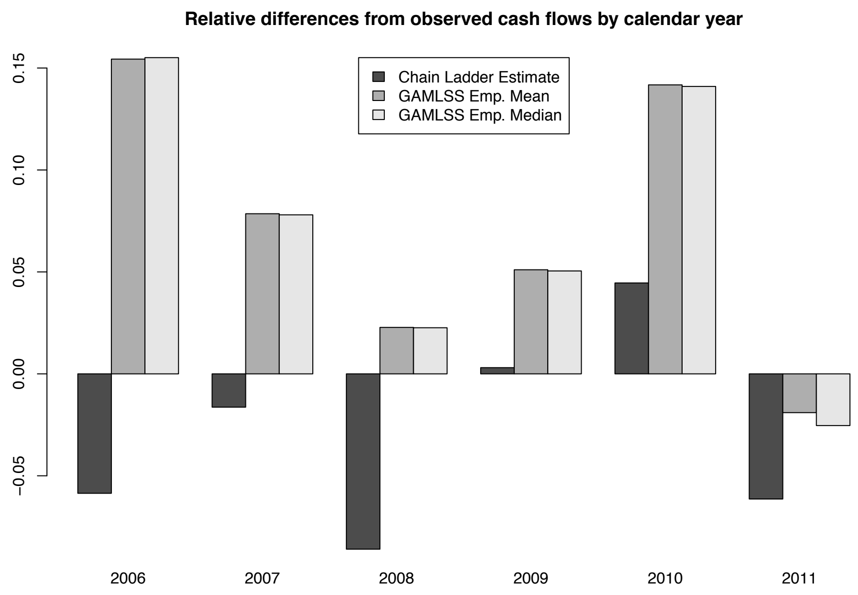


Figure B.2.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 2

Model	Distr.	Parameter	Polynomials	Factors	Interactions
2005	SEP1	μ	poly(DevLag,3), AYear		
		$\ln(\sigma)$	DevLag, poly(AYear,3)	DevLag 1,2, AYear 1991	
		ν	1		
		$\ln(\tau)$	1		
2006	SEP1	μ	poly(DevLag,3)		
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,3)	DevLag 3,4, AYear 1991,2000	
		ν	1		
		$\ln(\tau)$	1		
2007	SEP1	μ	poly(DevLag,3)		
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,3)	AYear 1991,2000, DevLag 3,4	
		ν	1		
		$\ln(\tau)$	1		
2008	SEP1	μ	poly(DevLag,3)	DevLag 3	
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,3)	AYear 1991,2000, CYear 2004,2008, DevLag 2	
		ν	1		
		$\ln(\tau)$	1		
2009	SEP1	μ	poly(DevLag,4)	DevLag 3	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,4)	AYear 1991,1993, DevLag 2	
		ν	1		
		$\ln(\tau)$	1		
2010	SEP1	μ	poly(DevLag,4), poly(AYear,3)	DevLag 3	
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,5)	DevLag 2	
		ν	1		
		$\ln(\tau)$	1		

Table B.2.1.: All six models for LoB 2

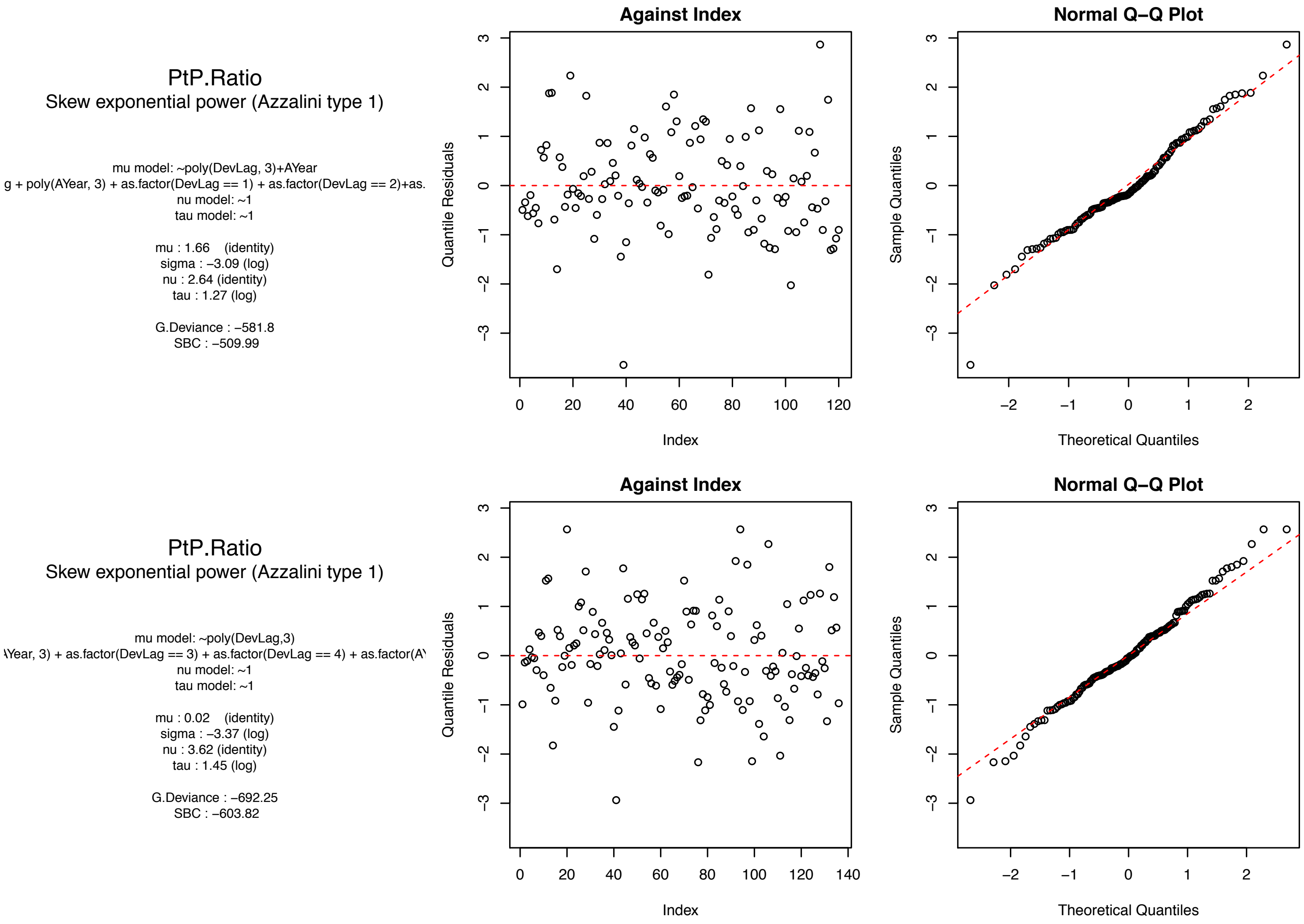


Figure B.2.2.: Residuals plots for LoB 2 (CYear 2005 & 2006)

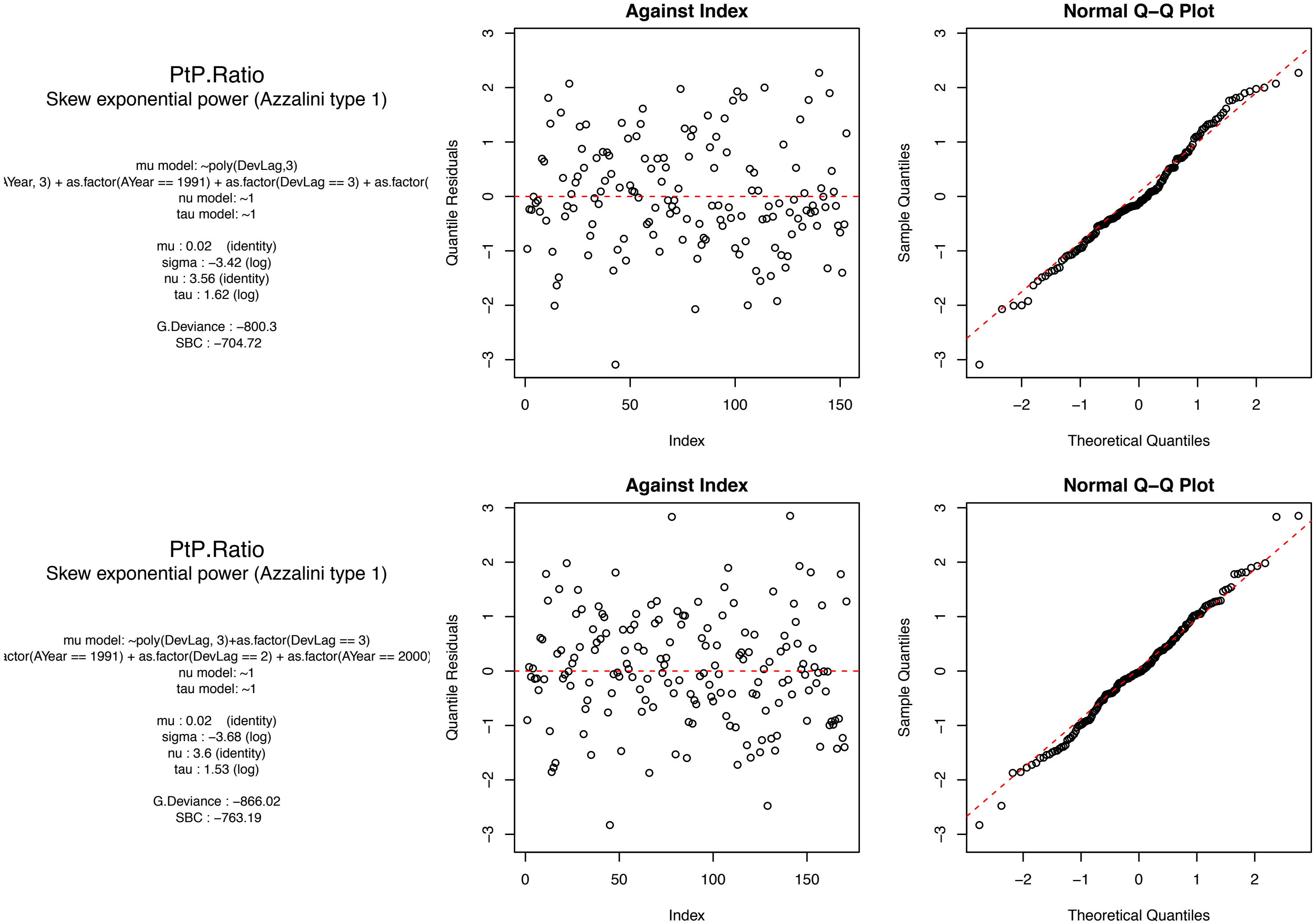


Figure B.2.3.: Residuals plots for LoB 2 (CYear 2007 & 2008)

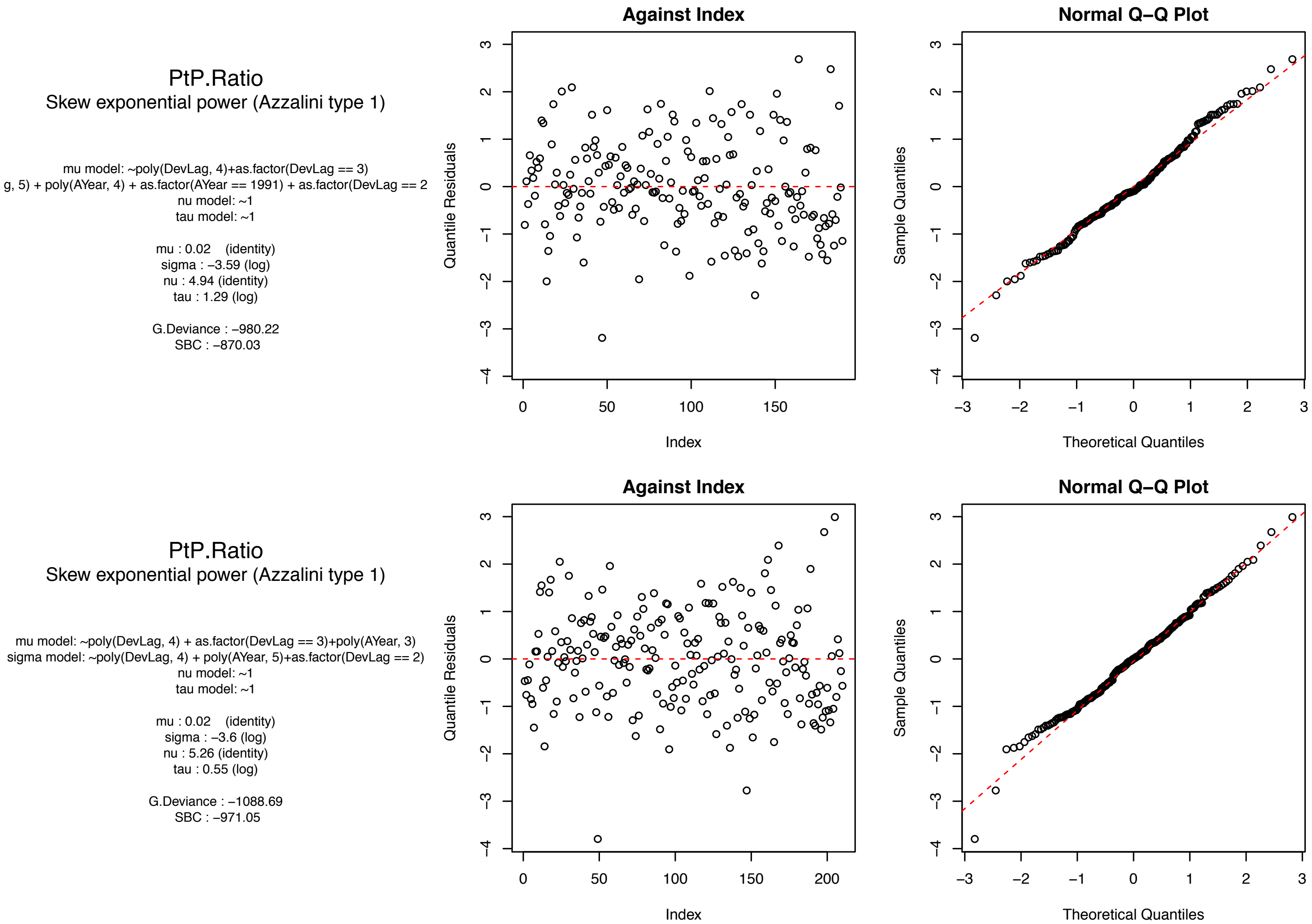


Figure B.2.4.: Residuals plots for LoB 2 (CYear 2009 & 2010)

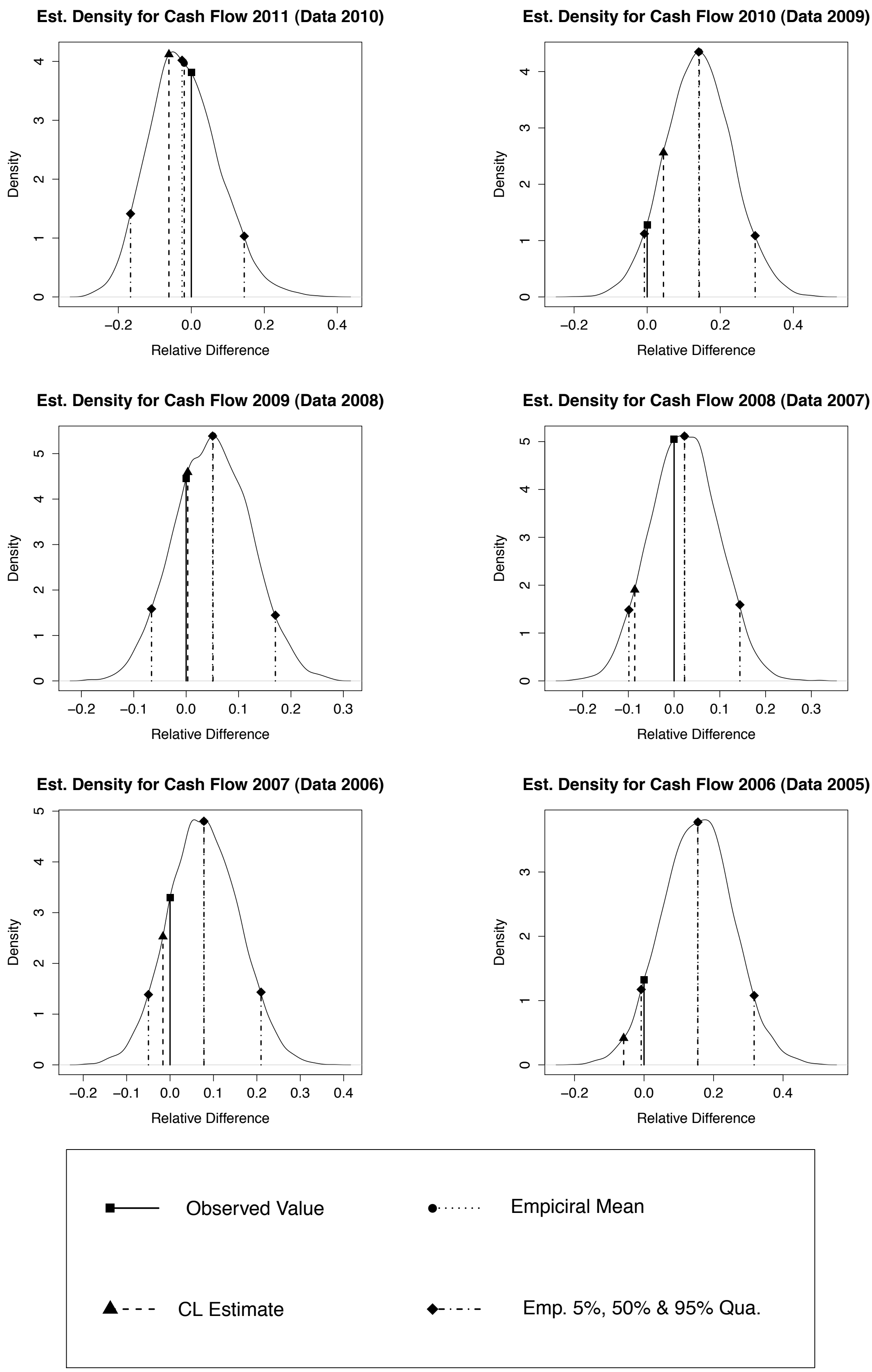
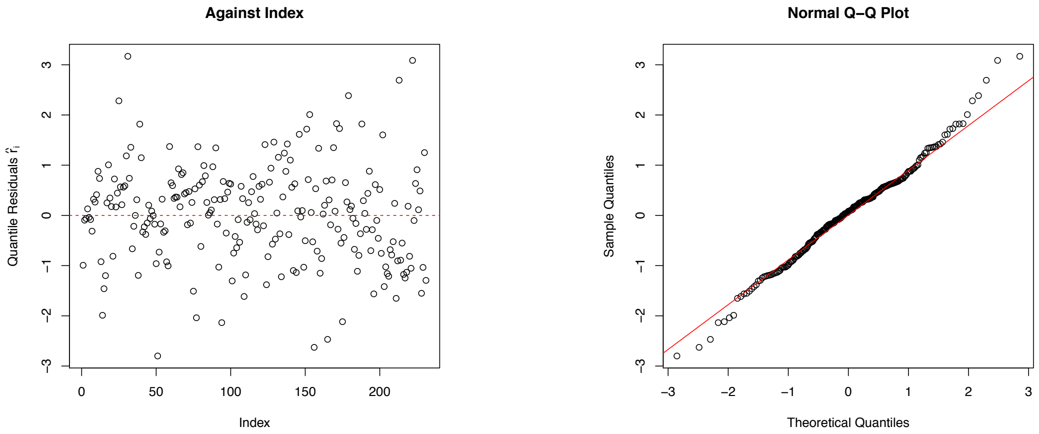


Figure B.2.5.: Predictive distributions for LoB 2

B.2.2. Latest Model & Reserving

The model using most recent data is

$$
\begin{aligned}
PtP.Ratio \sim \text{SEP1}\Big(\quad & \mu = DevLag + (DevLag = 1) + (DevLag = 2) \\
& \quad + (DevLag = 3) + (DevLag = 3) : poly(AYear, 2), \\
& \ln(\sigma) = poly(DevLag, 4) + poly(AYear, 2) + poly(CYear, 2) \\
& \quad + (AYear = 2002) + (DevLag = 3) + (CYear = 2008), \\
& \nu = 1, \\
& \ln(\tau) = 1 \Big).
\end{aligned}
$$



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure B.2.6.: Residual plots for Line of Business 2

Statistics of reserve distributions for lines of business 2:

	GAMLSS	CL
Mean	0.0000	-0.0649
Variance	0.0100	0.0158

Table B.2.2.: LoB 2 total reserve statistics

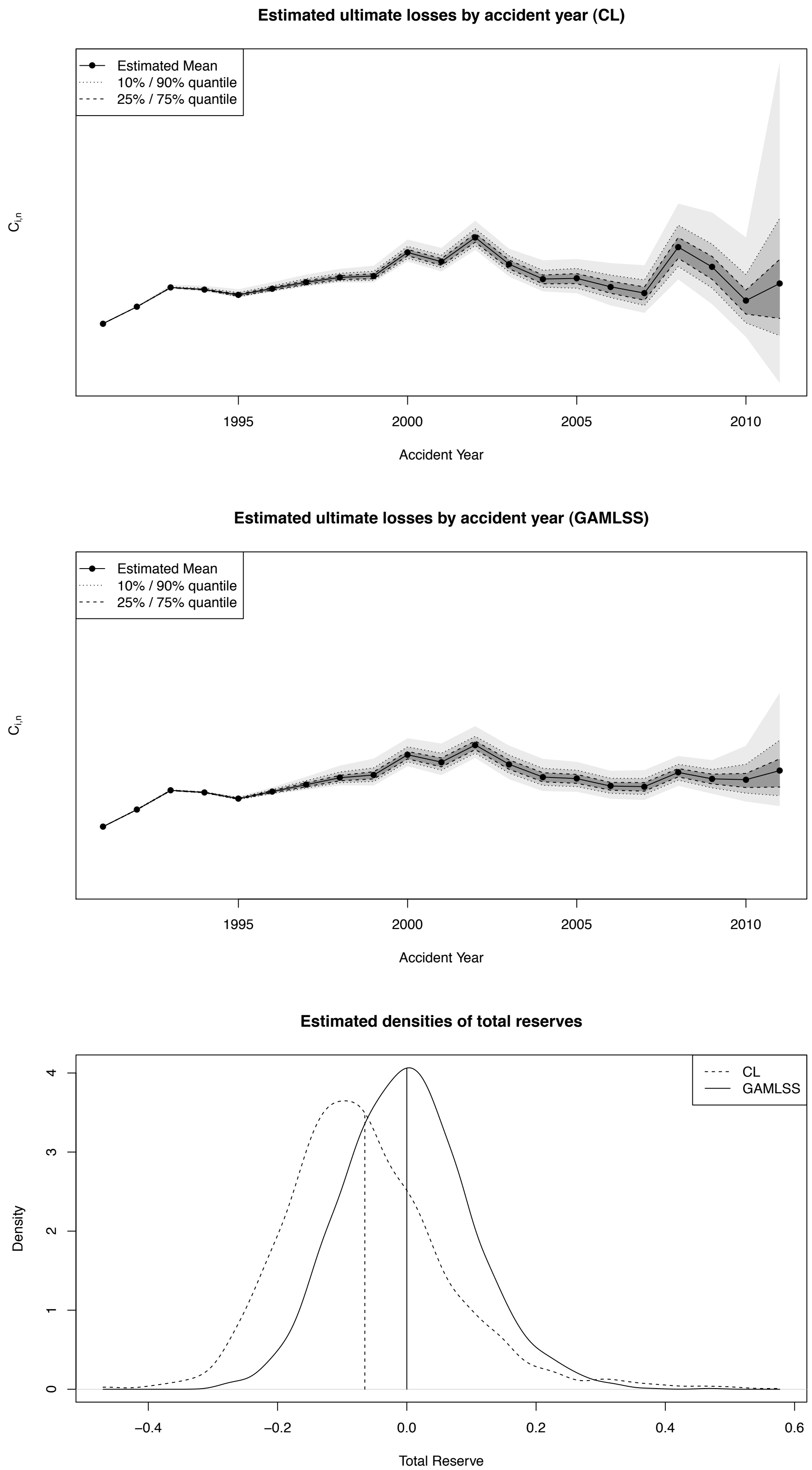


Figure B.2.7.: Ultimate loss estimates and estimated total reserve densities (LoB 2)

B.3. Line of Business 3

B.3.1. Cash Flow Analysis

Cash flow analysis (see Figure B.3.1) shows that chain ladder always underestimates cash flows. This situataion has been described in detail in Section 6.2.

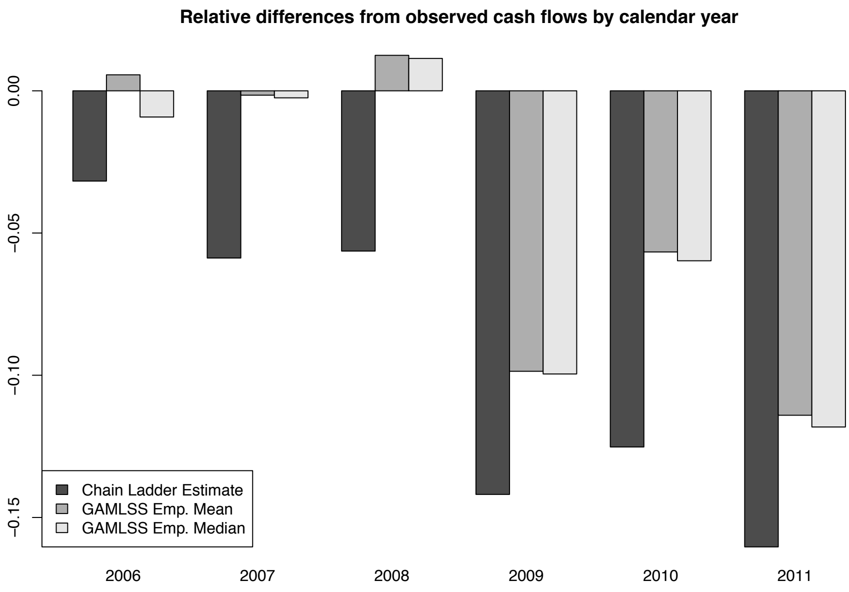


Figure B.3.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 3

Model	Distr.	Parameter	Polynomials	Factors	Interactions
2005	SN1	μ	poly(DevLag,4)	DevLag 1,2, AYear > 2003	
		$\ln(\sigma)$	poly(DevLag,3), poly(AYear,4)	DevLag 2, AYear,1994	
		ν	1		
		τ	2 (fixed)		
2006	SN1	μ	poly(DevLag,4)	DevLag 1)	(DevLag 1):AYear, (DevLag 2):AYear
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,4)		
		ν	1		
		$(\ln(\tau))$	2 (fixed)		
2007	SN1	μ	poly(DevLag,4)	DevLag 1,2, (DevLag 2 & AYear > 2004)	(DevLag 1):AYear
		$\ln(\sigma)$	poly(DevLag,4), poly(AYear,4)		
		ν	1		
		τ	2 (fixed)		
2008	SN1	μ	poly(DevLag,4)	DevLag 1,2,3,5	(DevLag 1):AYear, (DevLag 2):AYear
		$\ln(\sigma)$	poly(DevLag,3), poly(AYear,4)	AYear 1991	
		ν	1		
		τ	2 (fixed)		
2009	SN1	μ	poly(DevLag,4)	DevLag 1,2,3,5	(DevLag 1):AYear, (DevLag 2):AYear
		$\ln(\sigma)$	poly(DevLag,3), poly(AYear,4)	AYear 1991	
		ν	1		
		τ	2 (fixed)		
2010	SN1	μ	poly(DevLag,4)	DevLag 1,2,3,5	(DevLag 1):AYear, (DevLag 2):AYear
		$\ln(\sigma)$	poly(DevLag,3), poly(AYear,4)	AYear 1991	
		ν	1		
		τ	2 (fixed)		

Table B.3.1.: All six models for LoB 3

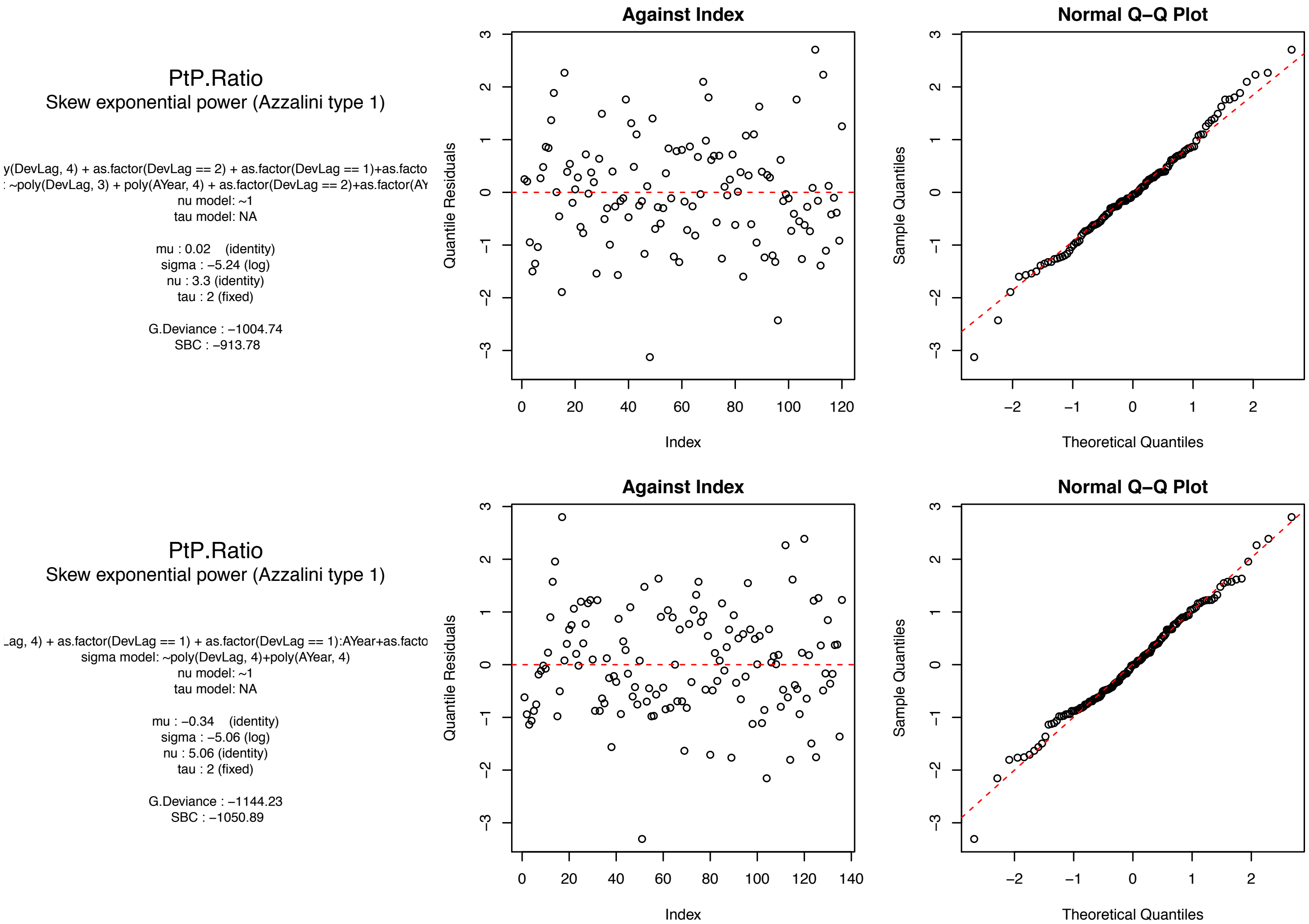


Figure B.3.2.: Residuals plots for LoB 3 (CYear 2005 & 2006)

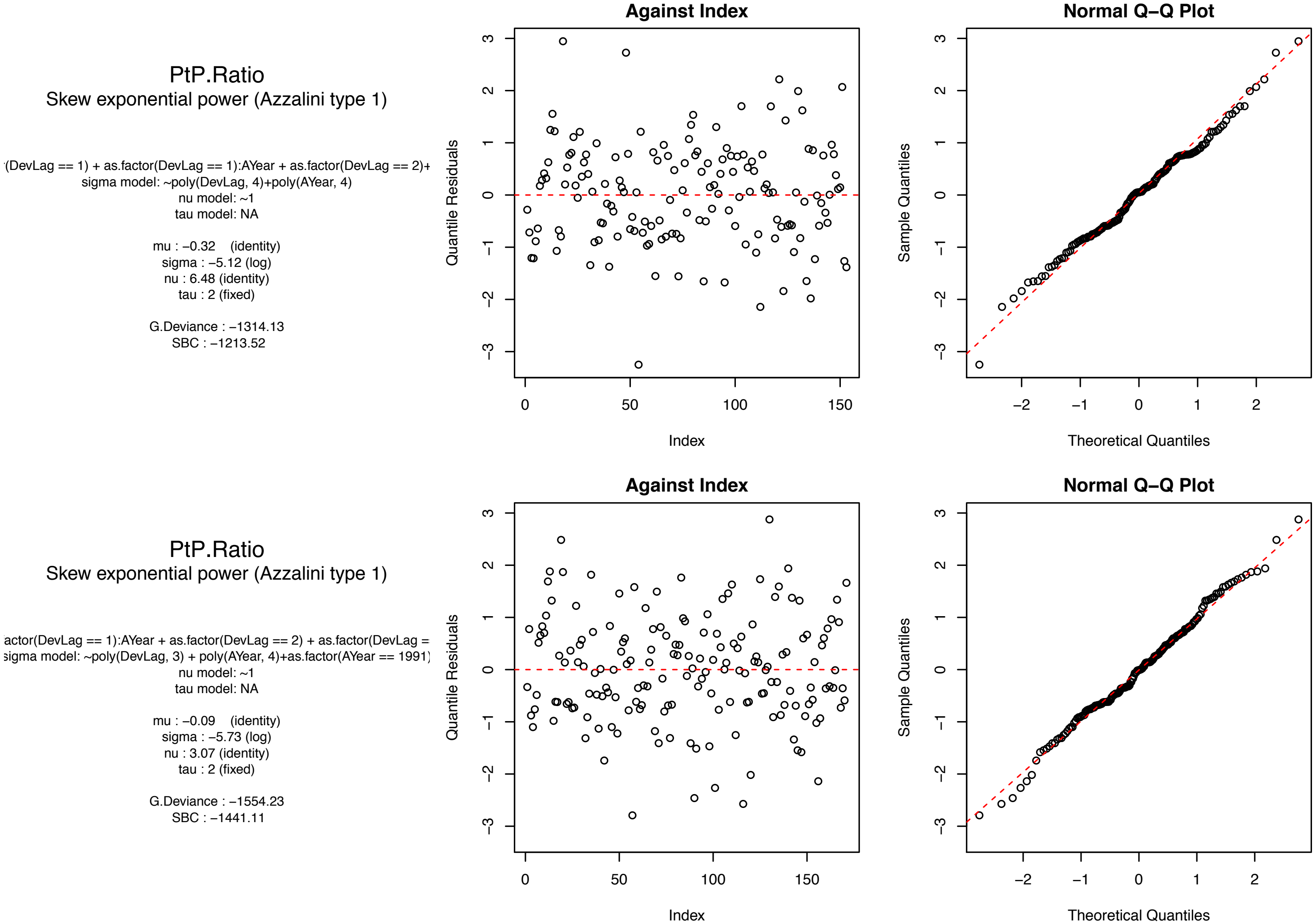


Figure B.3.3.: Residuals plots for LoB 3 (CYear 2007 & 2008)

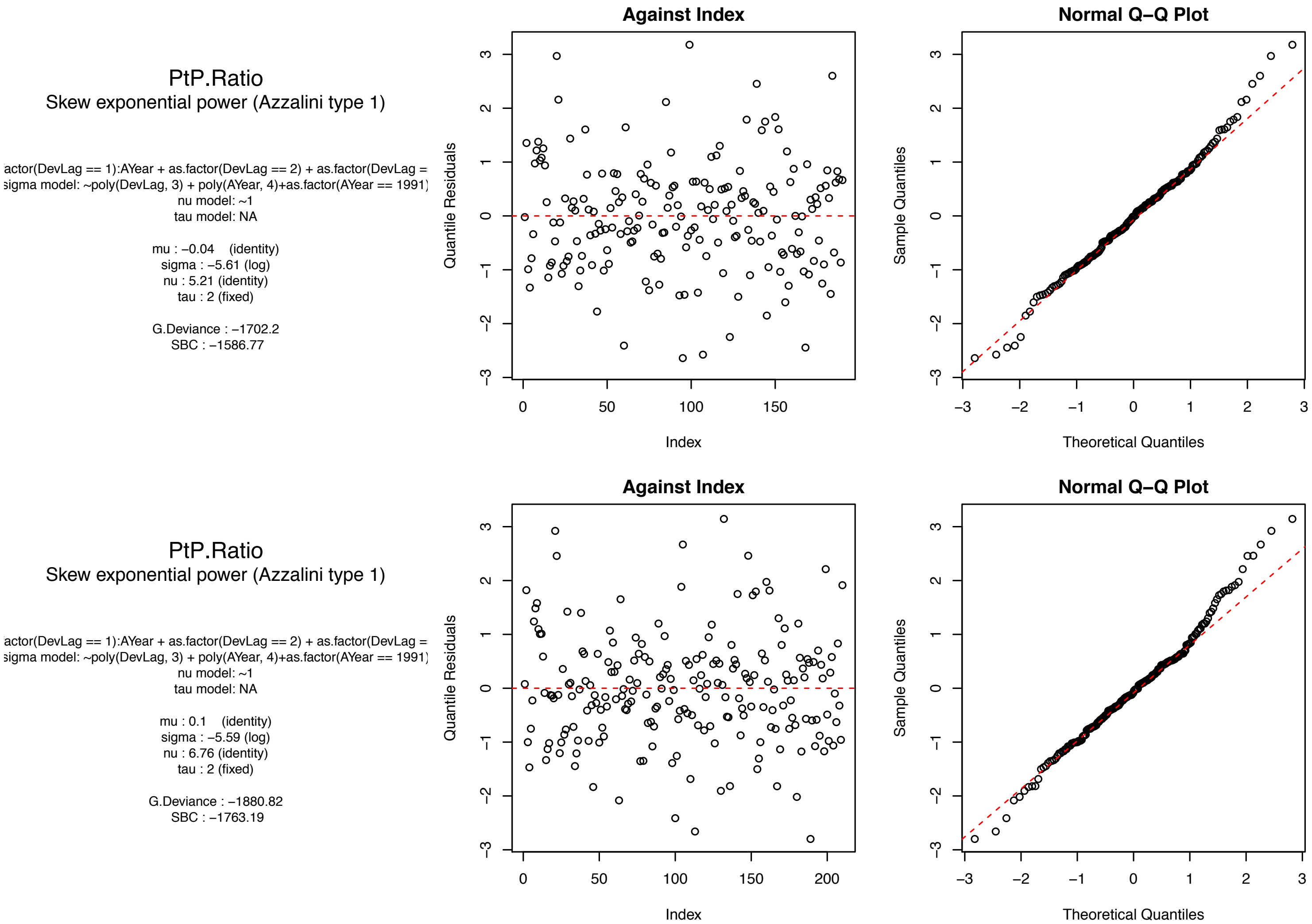


Figure B.3.4.: Residuals plots for LoB 3 (CYear 2009 & 2010)

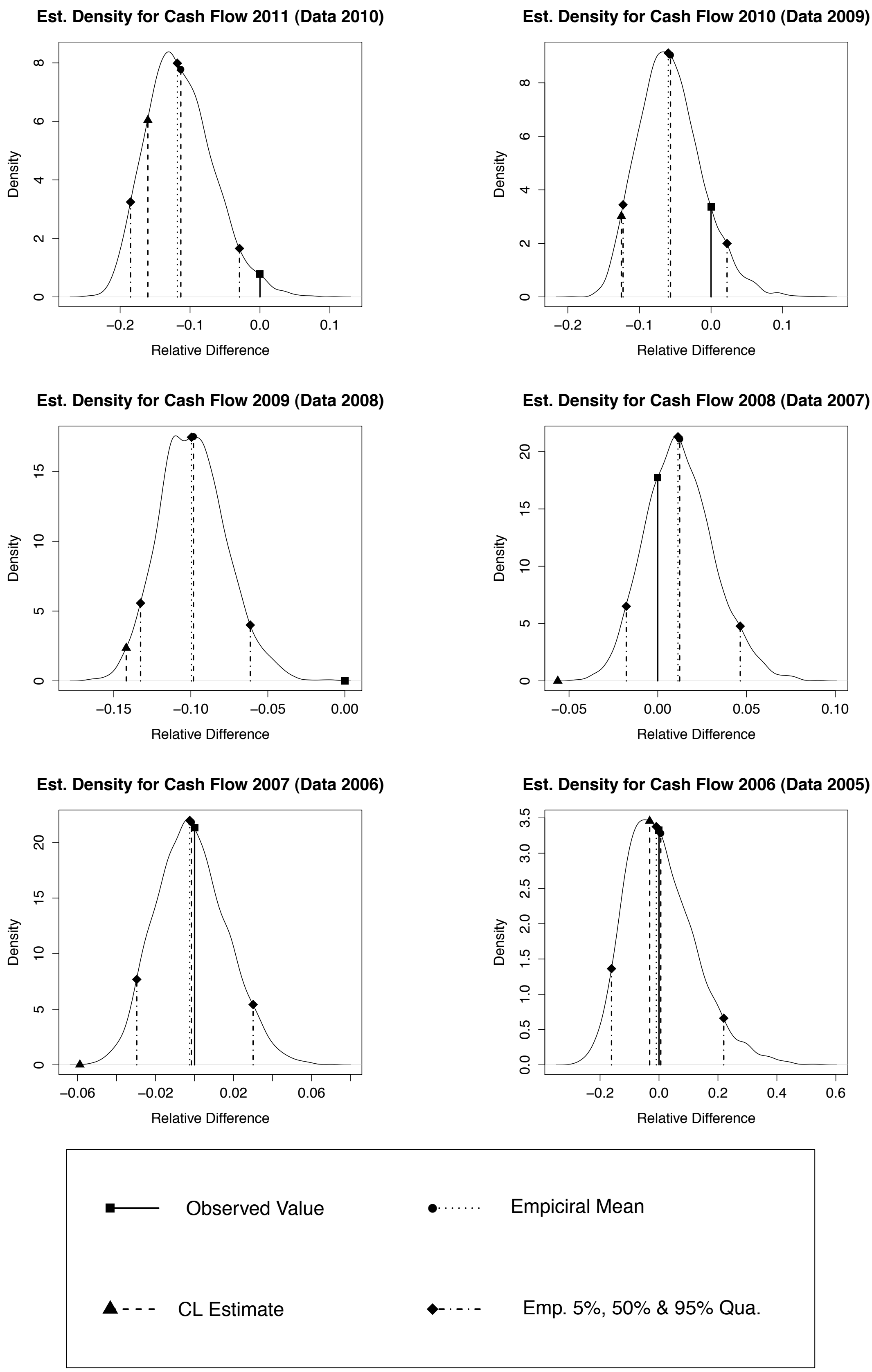
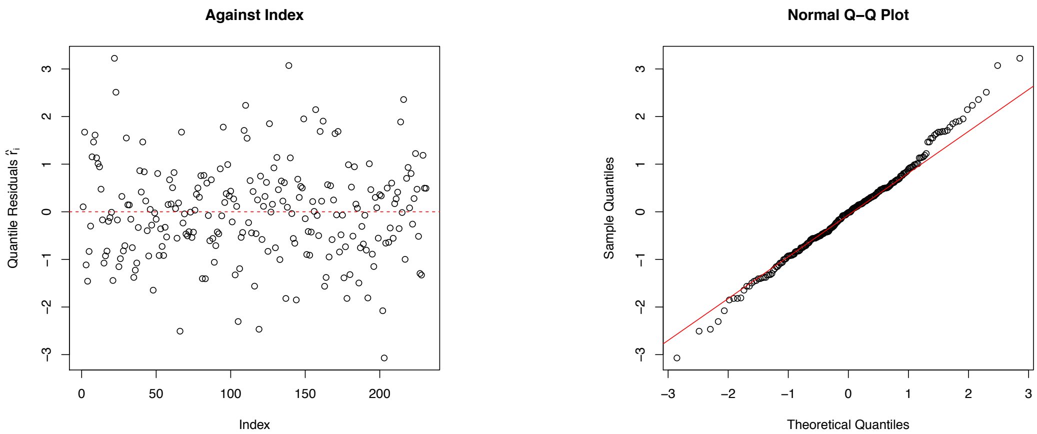


Figure B.3.5.: Predictive distributions for LoB 3

B.3.2. Latest Model & Reserving

The model using most recent data is

$$
\begin{aligned}
PtP.Ratio \sim \mathrm{SN1}\Big(\quad & \mu = poly(DevLag, 4) + (DevLag = 1) + (DevLag = 2) \\
& \quad + (DevLag = 3) + (DevLag = 5) \\
& \quad + DevLag = 1) : AYear + (DevLag = 2) : AYear, \\
& \ln(\sigma) = poly(DevLag, 3) + poly(AYear, 4) + (AYear = 1991), \\
& \nu = 1, \\
& \tau := 2(\text{fixed})\Big).
\end{aligned}
$$



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure B.3.6.: Residual Plots for Line of Business 3

Statistics of reserve distributions for lines of business 3:

	GAMLSS	CL
Mean	0.0000	-0.1152
Variance	0.0142	0.0030

Table B.3.2.: LoB 3 total reserve statistics

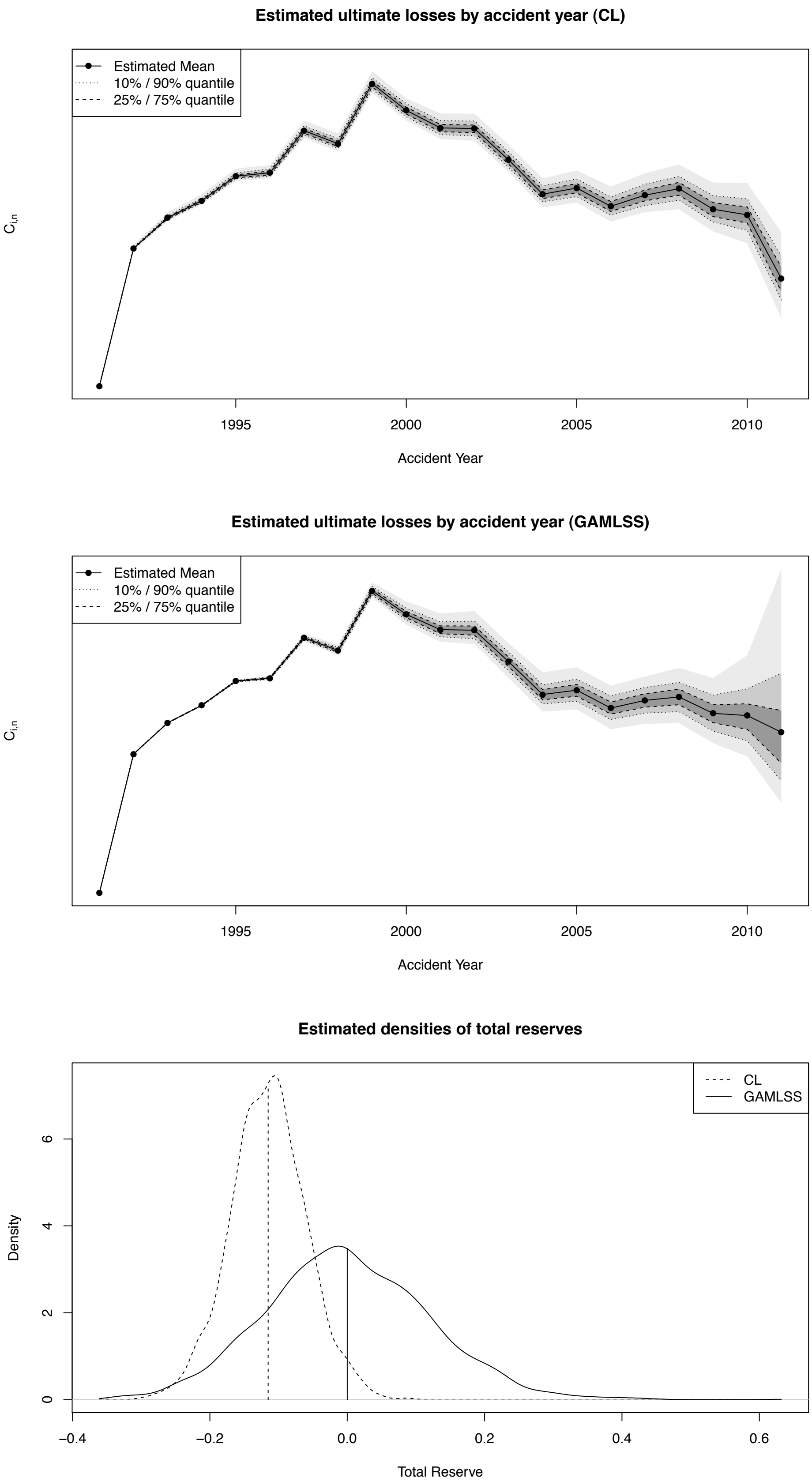


Figure B.3.7.: Ultimate loss estimates and estimated total reserve densities (LoB 3)

B.4. Line of Business 4

B.4.1. Cash Flow Analysis

Cash flow analysis shows that chain ladder over-estimates the cashflow for all years. The GAMLSS does so for four years and a big under-estimation for cash flow of calendar year 2010. The data shows no outlier here, so this seems to be an unusual estimate of the GAMLSS. So for this line of business without further research the chain ladder method might be the better choise because it is more constant in estimating the cash flows.

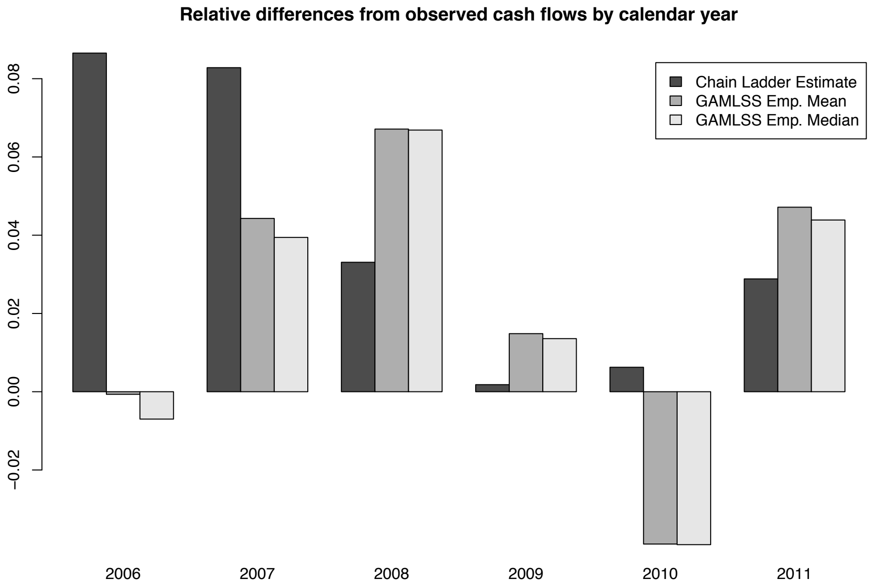


Figure B.4.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 4

Model	Distr.	Parameter	Polynomials	Factors	Interactions
2005	SEP1	μ	poly(DevLag,5), poly(AYear,3)	DevLag 2,3, CYear 2005	
		$\ln(\sigma)$	poly(DevLag,2)	AYear 1991	
		ν	1		
		τ	1.5 (fixed)		
2006	SEP1	μ	poly(DevLag,5), poly(AYear,3)	DevLag 2,3, CYear 2005,2006	
		$\ln(\sigma)$	poly(DevLag,2)	AYear 1992, DevLag 12	
		ν	1		
		τ	1.5 (fixed)		
2007	SEP1	μ	poly(DevLag,5), poly(AYear,3)	DevLag 2,3, CYear 2004	(DevLag 2):AYear, (DevLag 3):AYear
		$\ln(\sigma)$	poly(DevLag,4)	AYear 1991	
		ν	1		
		τ	4 (fixed)		
2008	SEP1	μ	poly(DevLag,5), poly(AYear,3)	DevLag 2,3	(DevLag 3):AYear
		$\ln(\sigma)$	poly(DevLag,2)	DevLag 2	(DevLag 4):AYear
		ν	1		
		τ	3 (fixed)		
2009	SEP1	μ	poly(DevLag,5), AYear	DevLag 2,3	(DevLag 2):AYear, (DevLag 3):AYear
		$\ln(\sigma)$	poly(DevLag,4), AYear	AYear 1991, CYear 2005	
		ν	1		
		τ	3.4 (fixed)		
2010	SEP1	μ	poly(DevLag,5), poly(AYear,3)	DevLag 2,3	(DevLag 2):AYear, (DevLag 3):AYear
		$\ln(\sigma)$	poly(DevLag,4)	DevLag 2, AYear 1991	
		ν	1		
		τ	2.5 (fixed)		

Table B.4.1.: All six models for LoB 4

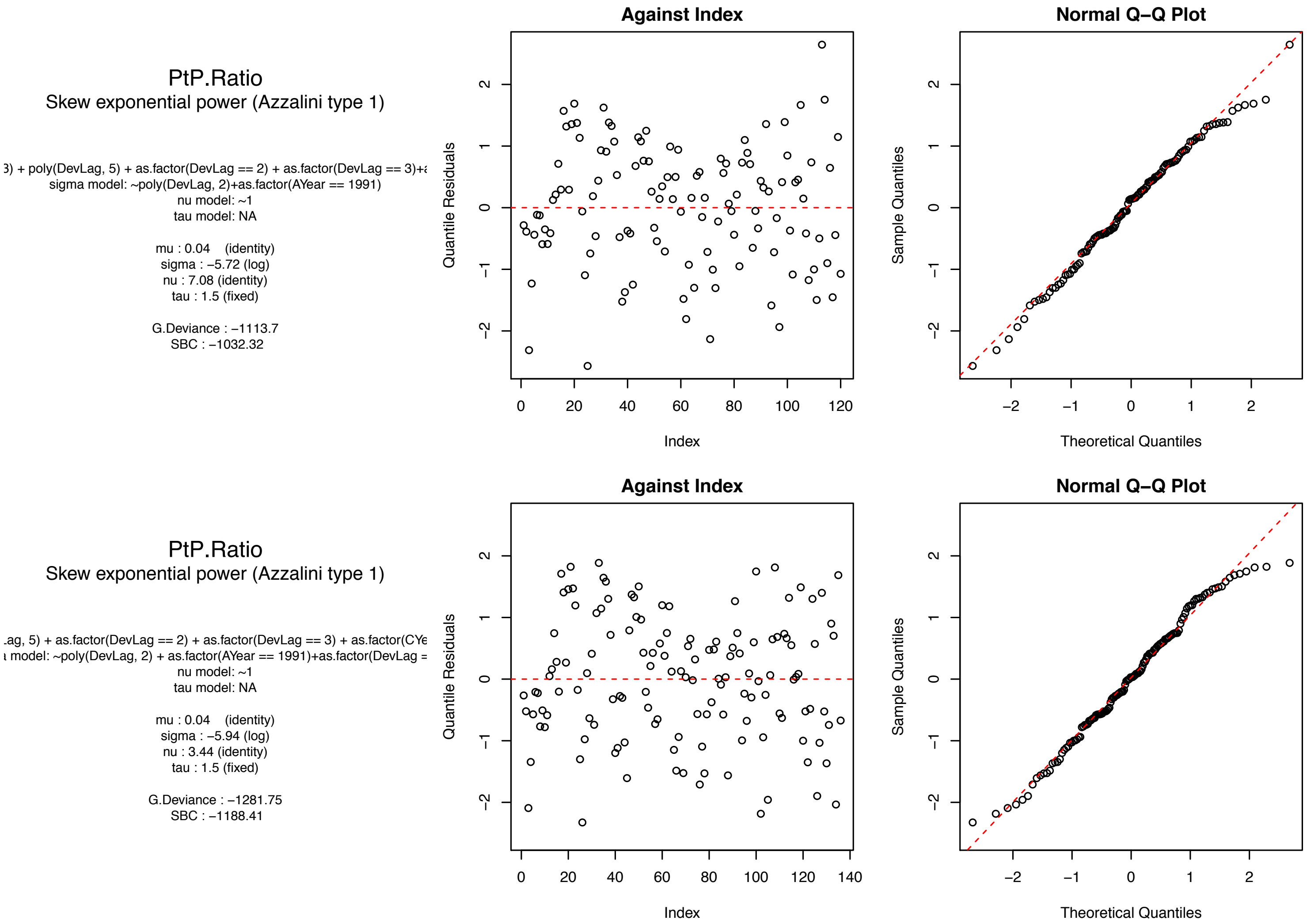


Figure B.4.2.: Residuals plots for LoB 4 (CYear 2005 & 2006)

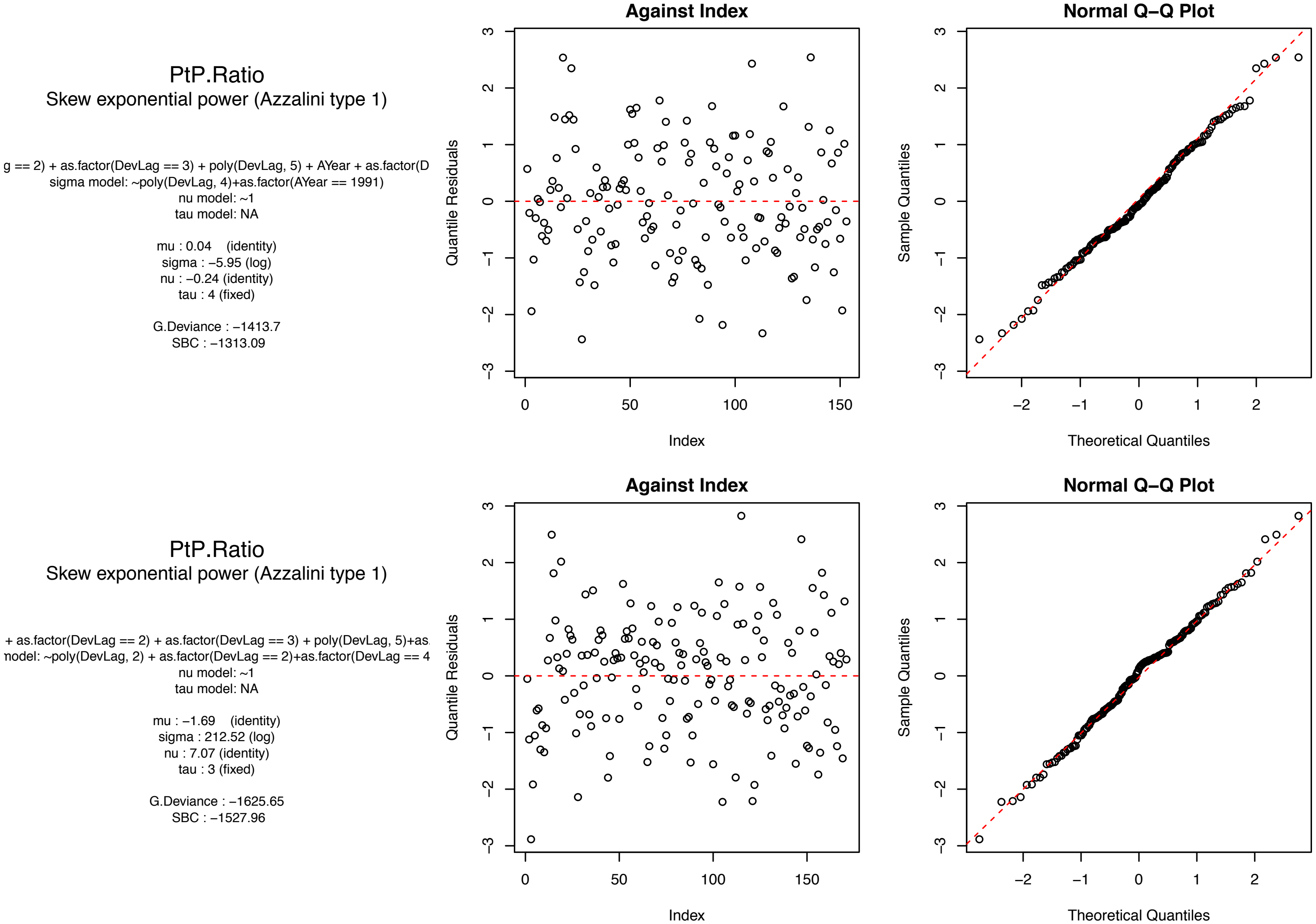


Figure B.4.3.: Residuals plots for LoB 4 (CYear 2007 & 2008)

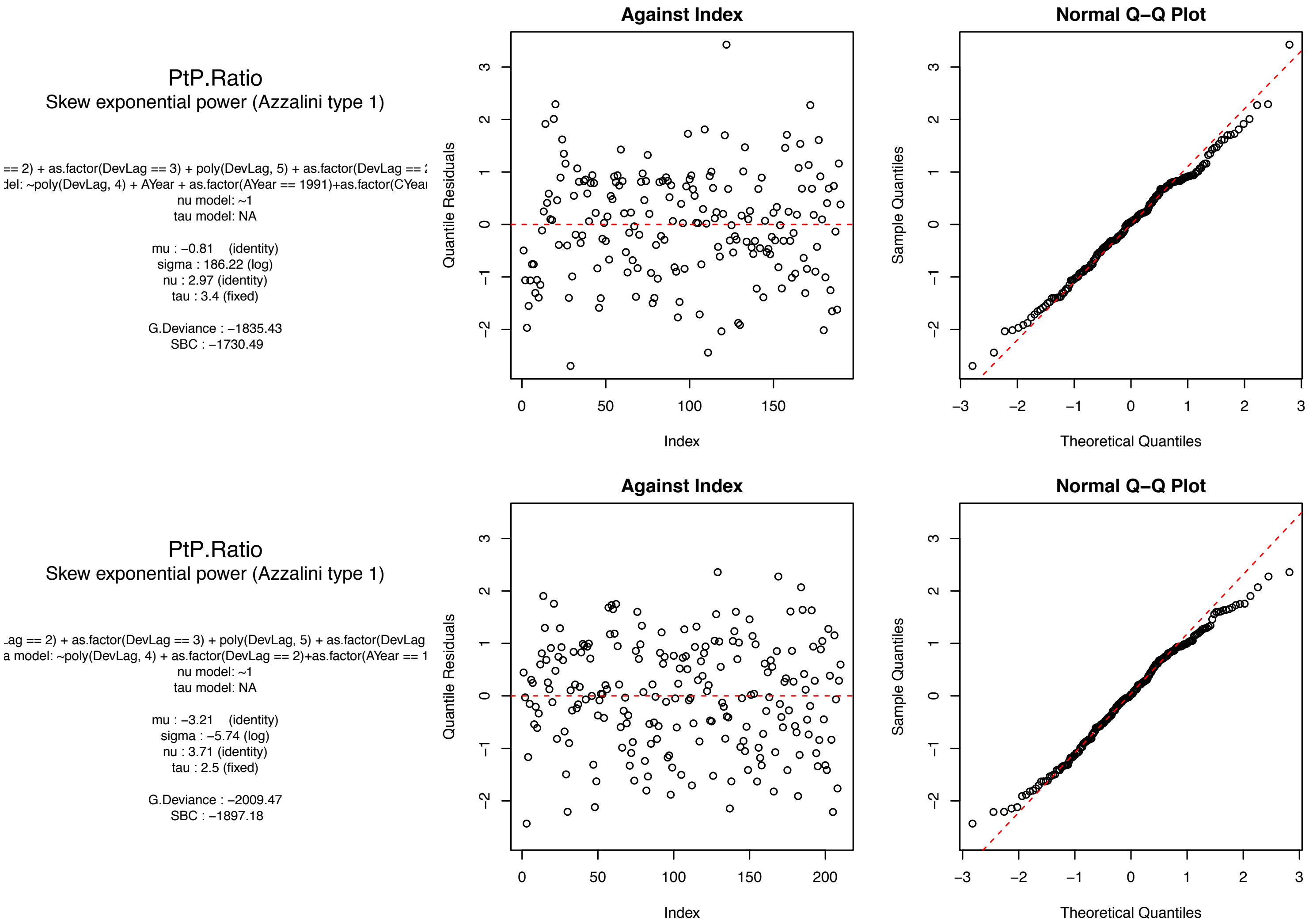


Figure B.4.4.: Residuals plots for LoB 4 (CYear 2009 & 2010)

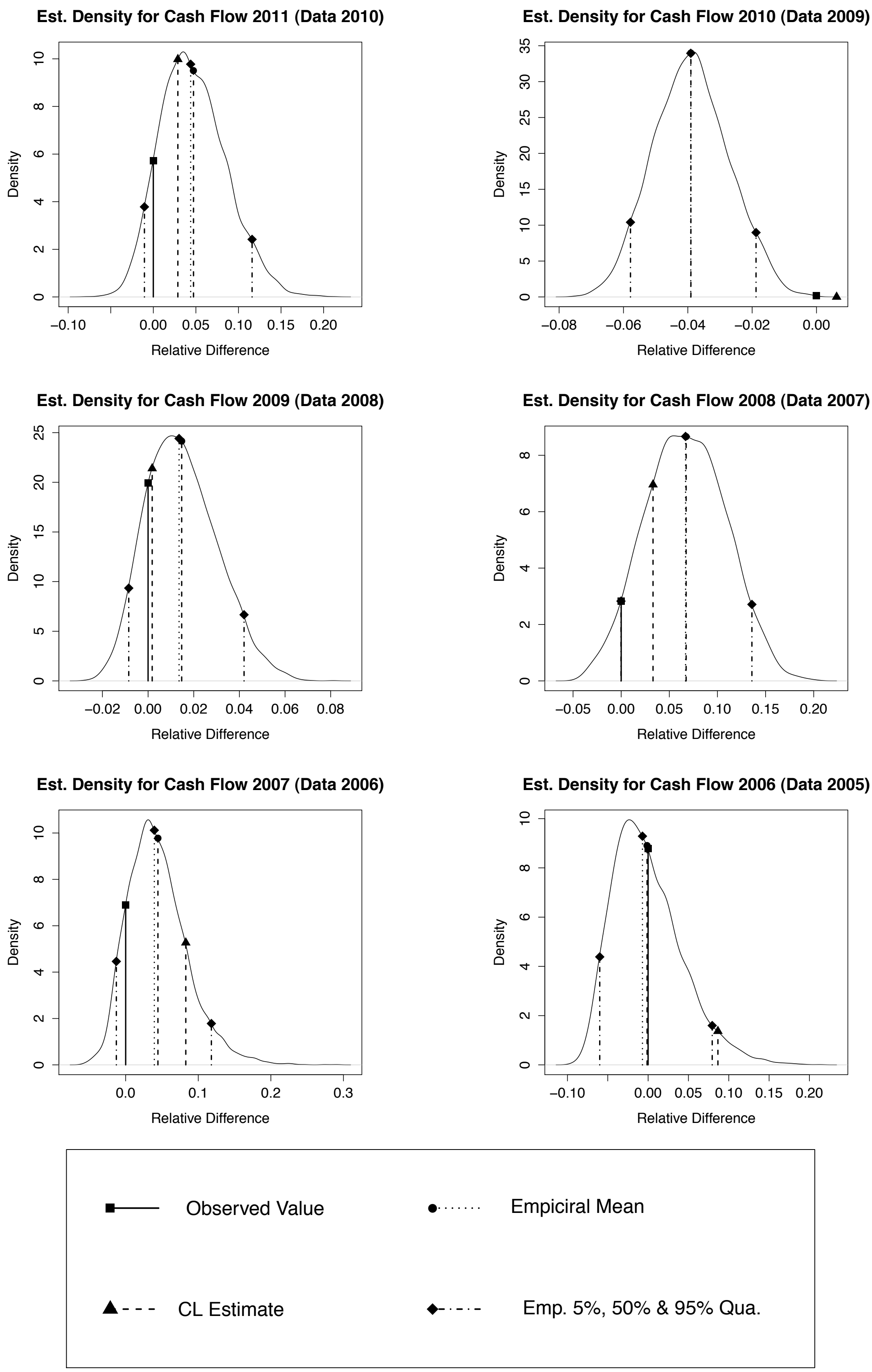
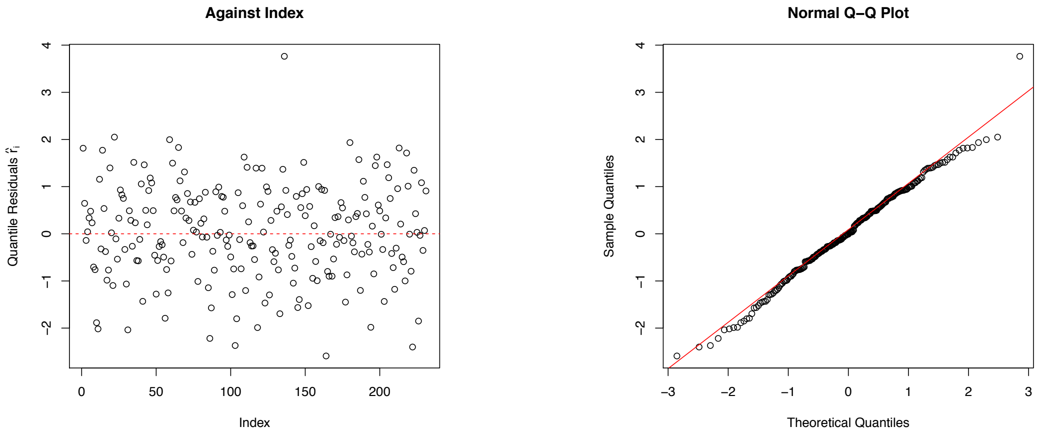


Figure B.4.5.: Predictive distributions for LoB 4

B.4.2. Latest Model & Reserving

The model using most recent data is

$$
\begin{aligned}
PtP.Ratio \sim \text{SEP1}\Big(\quad & \mu = poly(DevLag, 5) + AYear + (DevLag = 3) \\
& \qquad + (DevLag = 1) : poly(AYear, 2) + (DevLag = 2) : AYear, \\
\ln(\sigma) &= poly(DevLag, 3) + poly(AYear, 2) + (DevLag = 2) \\
& \qquad + (DevLag = 3) + (CYear = 2002), \\
& \nu = 1, \\
& \tau := 2.8(\text{fixed})\Big).
\end{aligned}
$$



(a) Plot of normalized quantile residuals against index

(b) Q-Q normal plot of normalized quantile residuals

Figure B.4.6.: Residual plots for Line of Business 4

Statistics of reserve distributions for lines of business 4:

	GAMLSS	CL
Mean	0.0000	-0.3309
Variance	0.0020	0.0006

Table B.4.2.: LoB 4 total reserve statistics

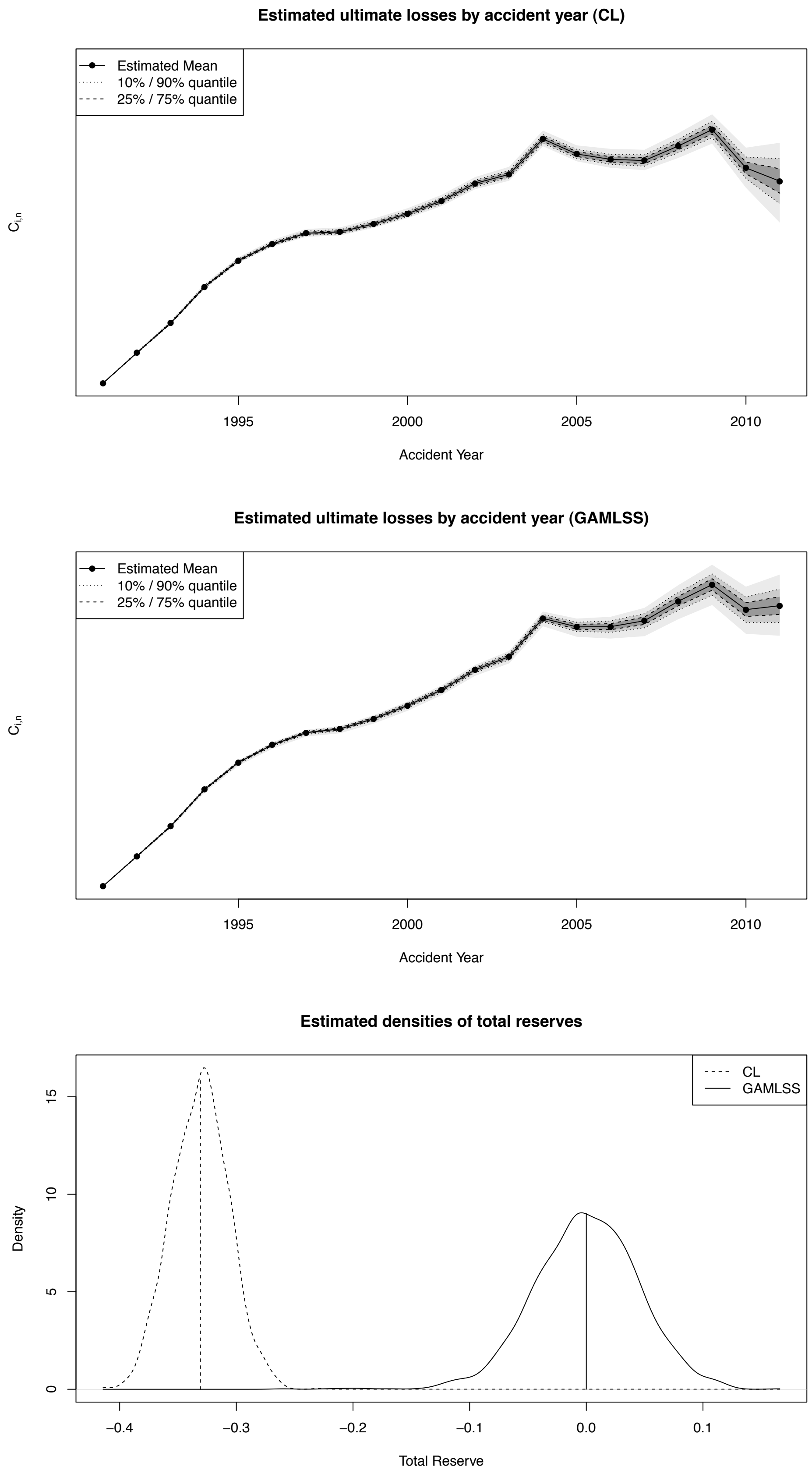


Figure B.4.7.: Ultimate loss estimates and estimated total reserve densities (LoB 4)

B.5. Line of Business 5

B.5.1. Cash Flow Analysis

Figure B.5.1 shows that except for the first cash flow chain ladder under-estimates all cash flows. The GAMLSS under-estimates the cash flows for all calendar years but not as strong as chain ladder. Data show that development factors increase for increasing accident years, causing an under-estimation when using the chain ladder factors. Thus we think the GAMLSS model is the better model in this case.

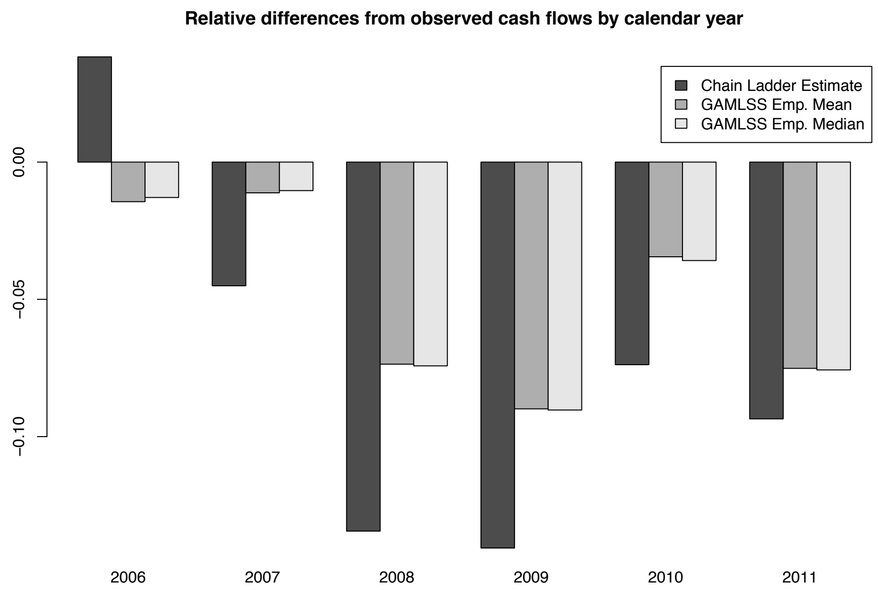


Figure B.5.1.: Chain ladder estimates, empirical means and empirical medians from GAMLSS models for LoB 5

Model	Distr.	Parameter	Polynomials	Factors	Interactions
2005	NO	μ	poly(DevLag,5), poly(AYear,4)	DevLag 1,2, AYear 1999	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2, CYear 2003, DevLag $>$ 6	
2006	NO	μ	poly(DevLag,5)	DevLag 1,2,3,7, AYear 1999, AYear $>$ 2002	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2, CYear 2003, DevLag $>$ 6	
2007	NO	μ	poly(DevLag,5)	DevLag 1,2,3,7	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2	
2008	NO	μ	poly(DevLag,5)	DevLag 1,2,3,7	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2	
2009	NO	μ	poly(DevLag,5)	DevLag 1,2,3,7	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2	
2010	NO	μ	poly(DevLag,5)	DevLag 1,2,3,7	
		$\ln(\sigma)$	poly(DevLag,5), poly(AYear,2)	DevLag 2	

Table B.5.1.: All six models for LoB 5.

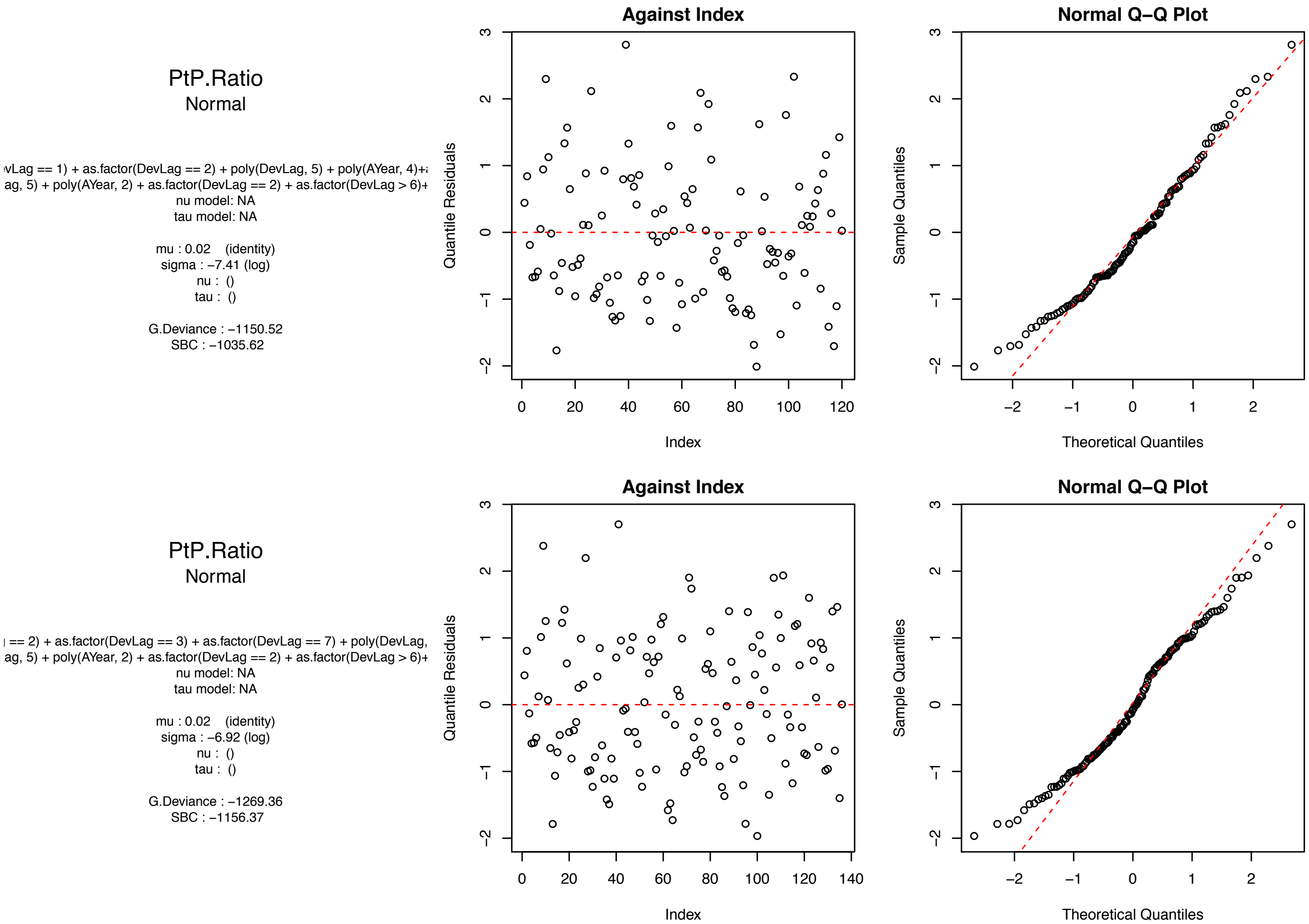


Figure B.5.2.: Residuals plots for LoB 5 (CYear 2005 & 2006)

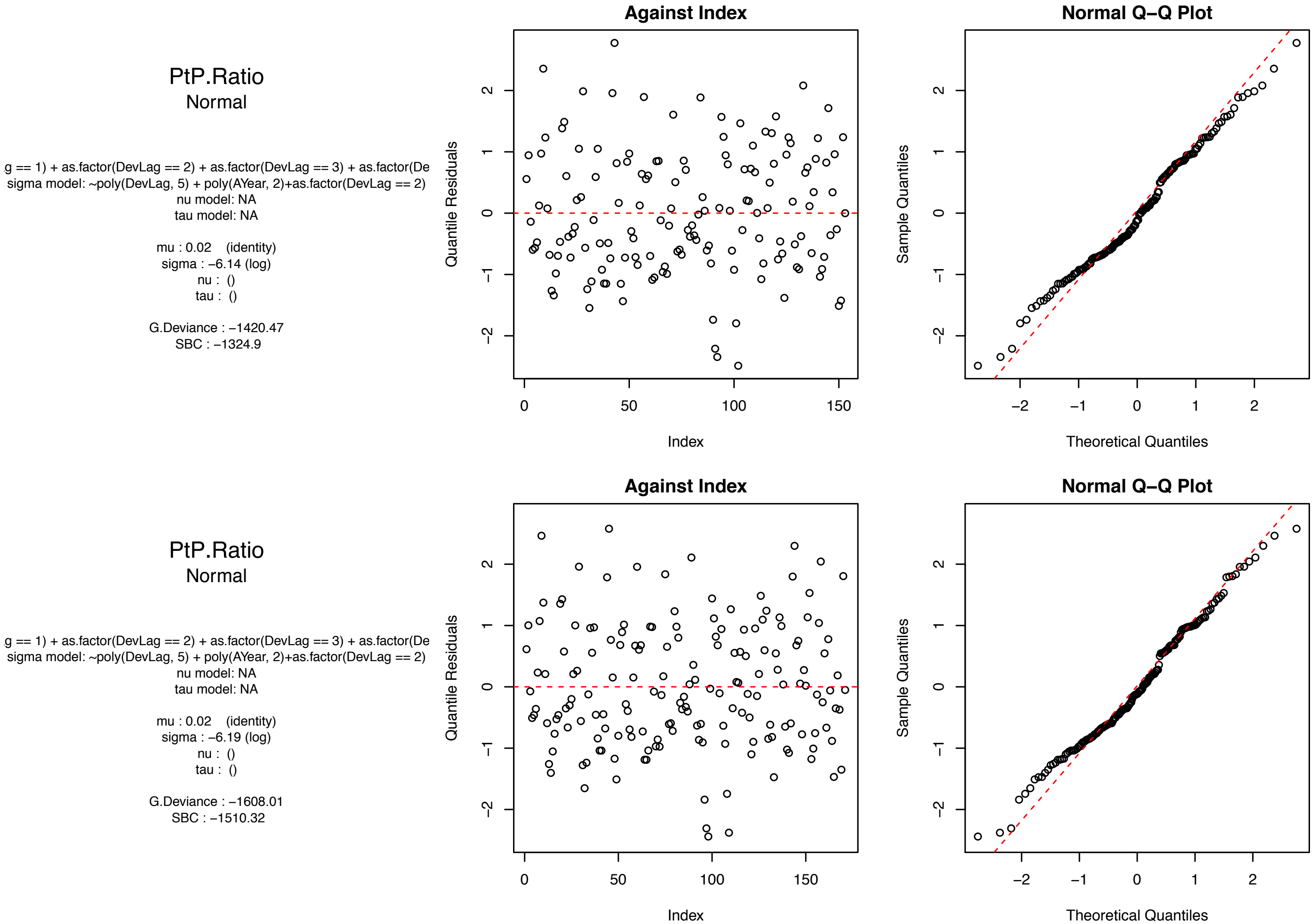


Figure B.5.3.: Residuals plots for LoB 5 (CYear 2007 & 2008)

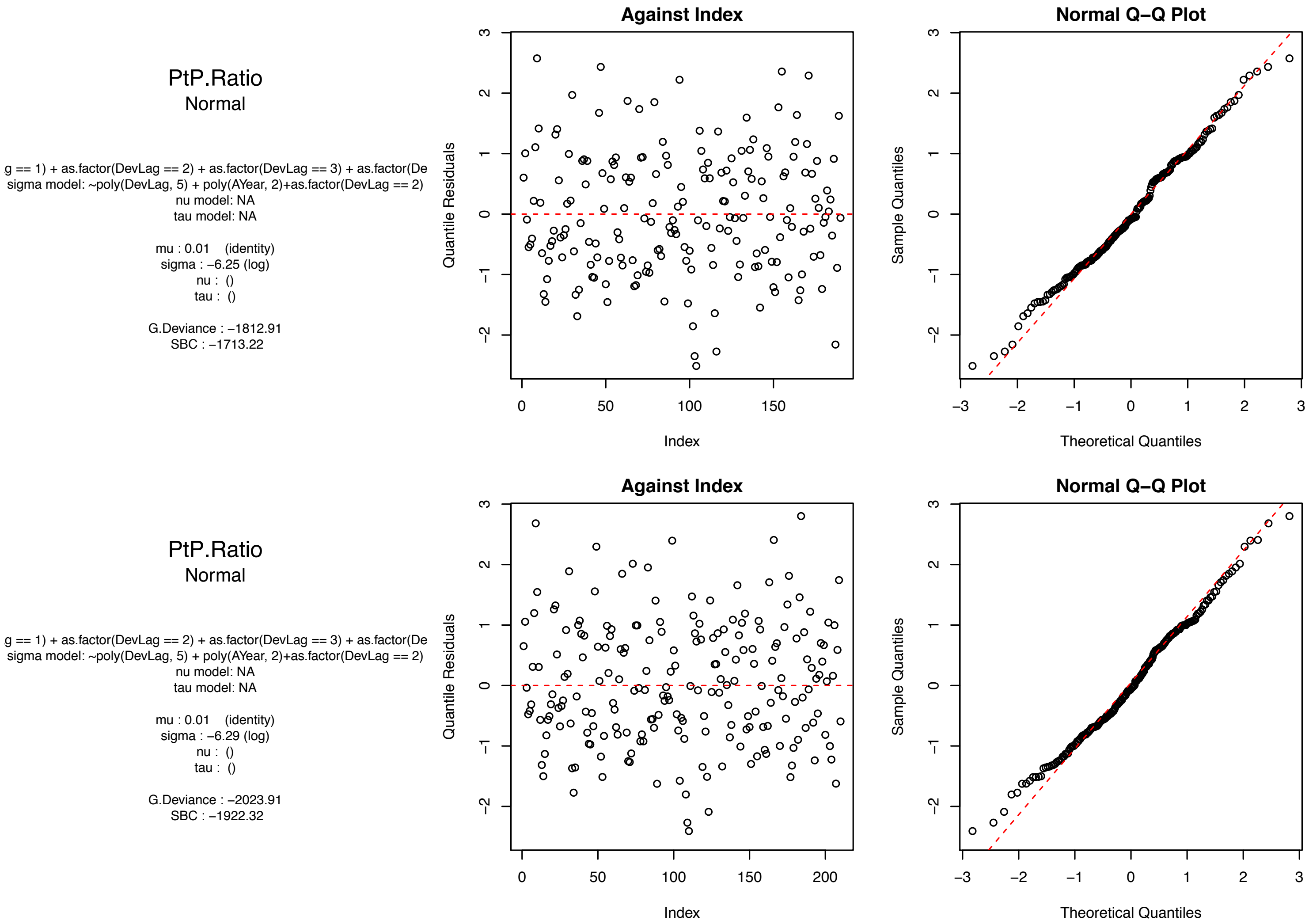


Figure B.5.4.: Residuals plots for LoB 5 (CYear 2009 & 2010)

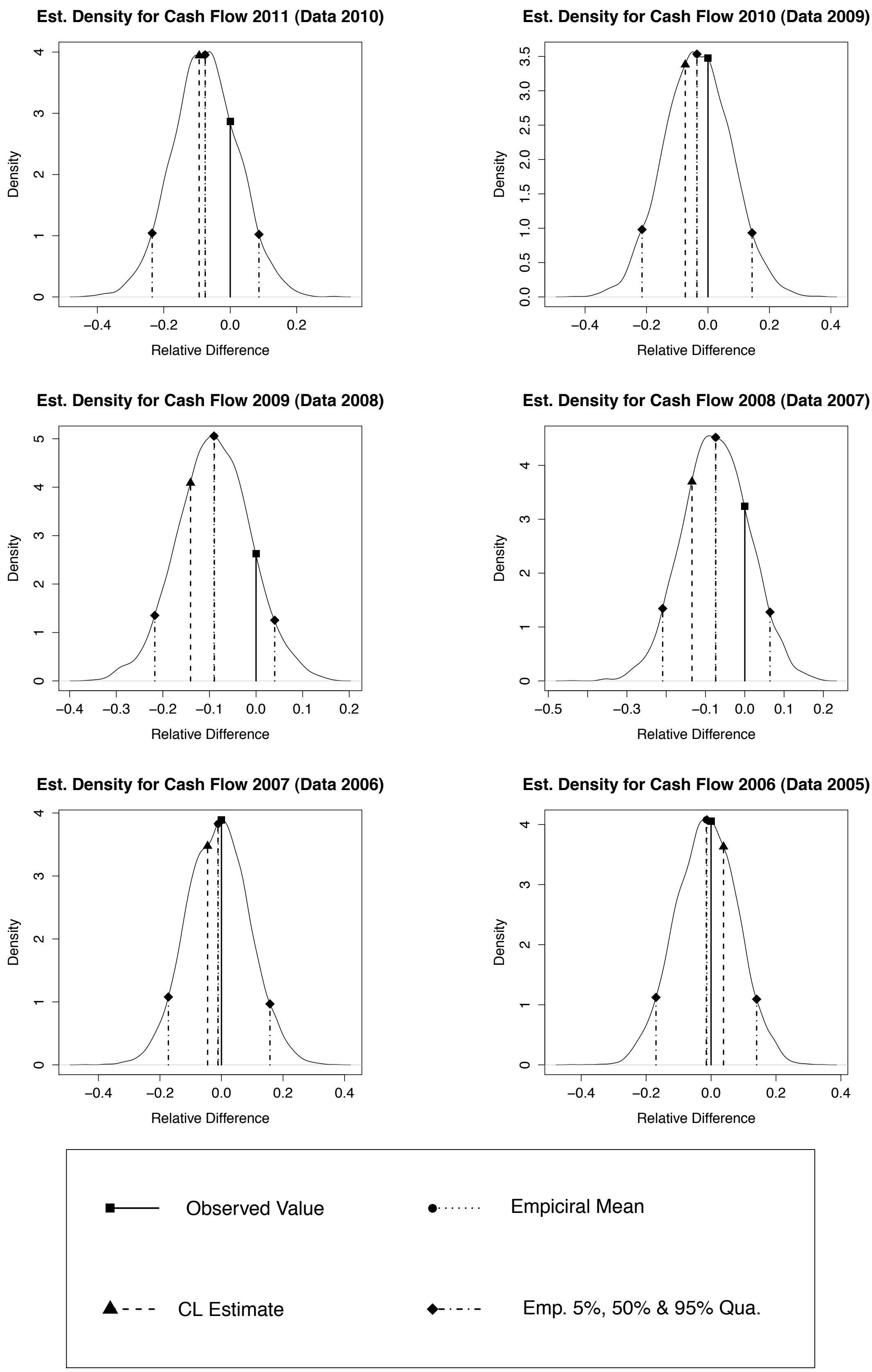
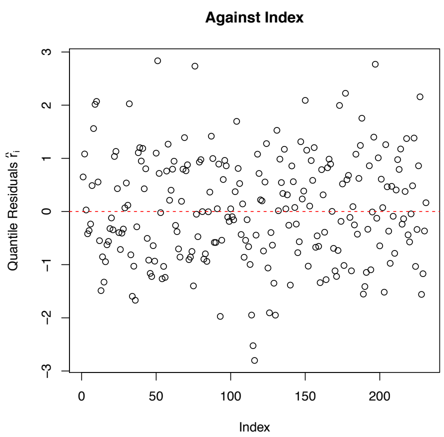


Figure B.5.5.: Predictive distributions for LoB 5

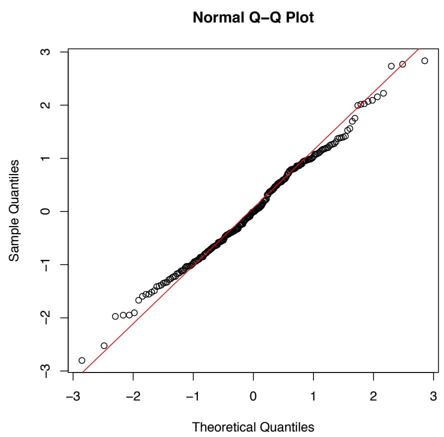
B.5.2. Latest Model & Reserving

The model using most recent data is

$$
\begin{aligned}
PtP.Ratio \sim \mathrm{NO}\Big(\quad & \mu = poly(DevLag, 5) + (DevLag = 1) + (DevLag = 2) \\
& \quad + (DevLag = 3) + (AYear = 2003), \\
& \ln(\sigma) = poly(DevLag, 5) + poly(AYear, 2) + (DevLag = 2) \\
& \quad + (AYear = 1999) + (AYear = 2004)\Big).
\end{aligned}
$$



(a) Plot of normalized quantile residuals against index



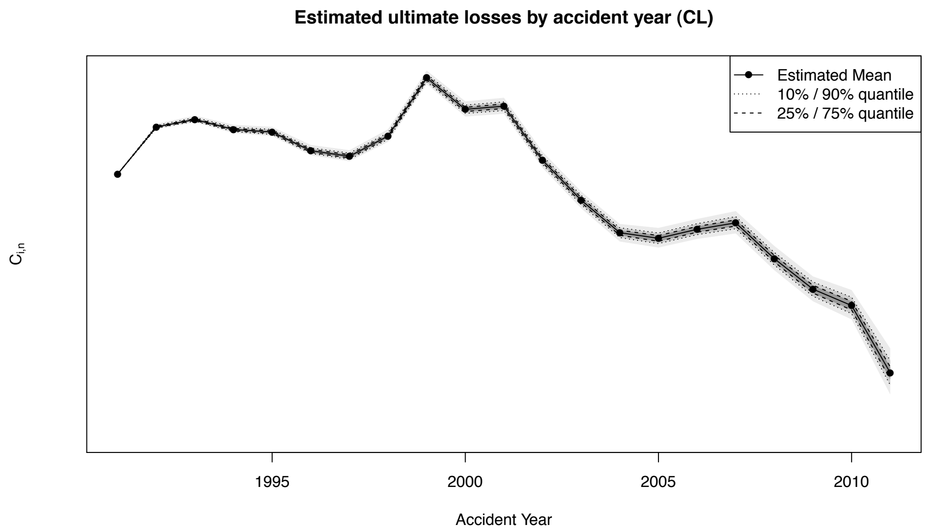
(b) Q-Q normal plot of normalized quantile residuals

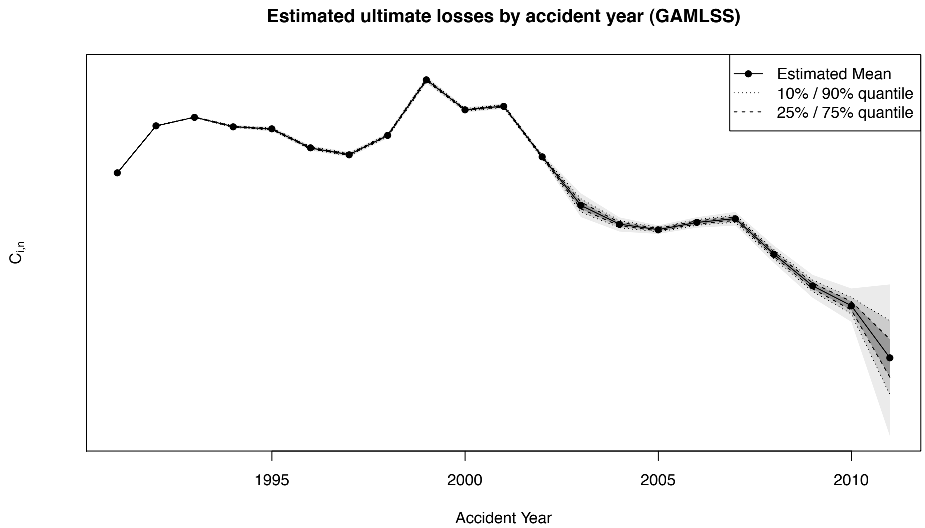
Figure B.5.6.: Residual plots for Line of Business 5

Statistics of reserve distributions for lines of business 5:

	GAMLSS	CL
Mean	0.0000	-0.0380
Variance	0.0014	0.0009

Table B.5.2.: LoB 5 total reserve statistics





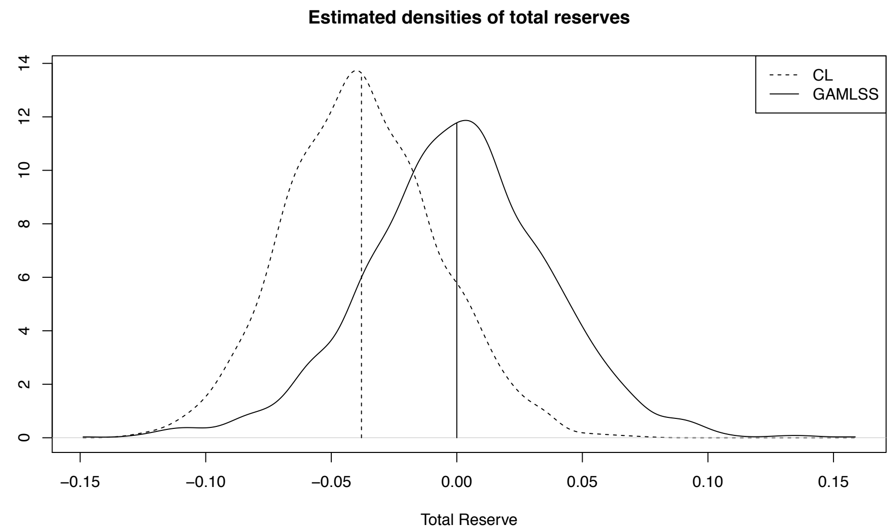


Figure B.5.7.: Ultimate loss estimates and estimated total reserve densities (LoB 5)

Index

accident year, 4
age-to-age factors, 9

calendar year, 4
cash flow, 77
chain ladder factor, 13
chain ladder factors, 9
concordant, 102
correlation
  Kendall's tau, 78
  Pearson's rho, 78

development lag, 4
discordant, 102
distribution
  (standardized) t, 35
  Gaussian, 34
  non-standardized t, 36
  power exponential type 2, 39
  skew exponential power type 1, 40
  skew normal, 41
  zero-inflated t, 38
  zero-inflated Gaussian, 37

Generalized Additive Model for Location, Scale and Shape (GAMLSS), 32
generalized linear model, 16
global deviance, 55

IBNER reserve, 6
IBNR reserve, 6

linear predictor, 16
link function, 16
loss triangles, 5

multiplicative approach, 8

normalized quantile residuals, 50

outstanding claims, 6

paid claims, 5
paid-to-premium ratio, 44

reserve, 3
  total, 7

Schwarz Bayesian criterion, 55

tied, 102

ultimate loss, 7

variance function, 17

Bibliography

A. Azzalini. Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199 – 208, 1986.

C. Braun. The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin*, 34(2), 2004.

P. D. England and R. J. Verrall. Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8, 2002.

P.D. England. Addendum to 'Analytic and bootstrap estimates of prediction errors in claims reserving'. *Actuarial Research Paper*, 138, 2001.

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

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Monographs on Statistics and Applied Probability. Taylor & Francis, 1990. ISBN 9780412343902.

S. Jackman. *Bayesian Analysis for the Social Sciences*. John Wiley and Sons, Ltd, 2009. ISBN 9780470686621.

E. Kremer. IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*, (1), 1982.

T. Mack. Distribution-free calculation of the standard error of the chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213 – 225, 1993.

T. Mack. Which stochastic model is underlying the chain ladder method? *IME*, 15:133 – 138, 1994.

T. Mack. *Schadenversicherungsmathematik*. Number 28 in Angewandte Versicherungsmathematik. Verl. Versicherungswirtschaft, 2002. ISBN 388487957X.

T. Mack and G. Venter. A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: Mathematics and Economics*, 26:101 – 107, 2000.

P. McCullagh and J. A. Nelder. *Generalized Linear Models, Second Edition*. Monographs on Statistics and Applied Probability. Taylor & Francis, 1989.

M. Merz and M. Wüthrich. Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal*, 12(2), 2008.

M. Merz and M. Wüthrich. Combining chain-ladder and additive loss reserving methods for dependent lines of business. *Variance*, 3(2):270–291, 2009.

A. O'Hagan and T. Leonard. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1):201–203, 1976. doi: 10.1093/biomet/63.1.201.

A. E. Renshaw and R. J. Verrall. A stochastic model underlying the chain-ladder technique. *Proceedings of the ASTIN Colloquium in Cannes*, pages 45 – 69, 1994.

B. A. Rigby and D. M. Stasinopoulos. The gamlss project: a flexible approach to statistical modelling. *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling*, 2001.

B. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape. *Appl. Statist.*, 54, 2005.

P. Shi and E. W. Frees. Dependent loss reserving using copulas. *ASTIN Bulletin*, 54, 2011.

D. M. Stasinopoulos, B. A. Rigby, and C. Akantziliotou. *Instructions on how to use the gamlss package in R – Second Edition*, 2008. http://studweb.north.londonmet.ac.uk/ stasinom/papers/gamlss-manual.pdf.

R. J. Verrall. An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26:91 – 99, 2000.