Institute and Faculty of Actuaries

C01: Regression Models Based on Log-incremental Payments

Markus Gesmann

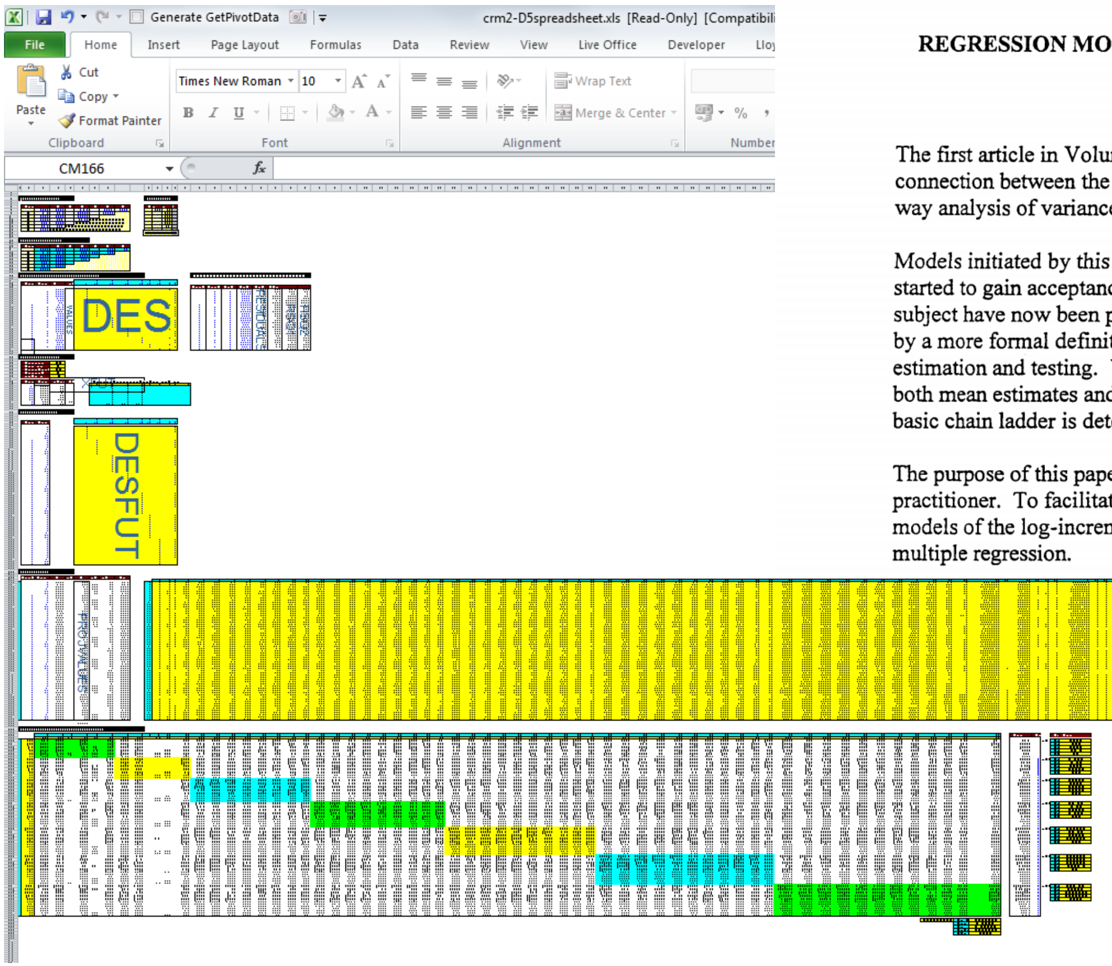
GIRO, 21 October 2015

What I'd like to cover today

- Go through ideas of the Claims Reserving Manual D5 (1997)
  - Assume claims follow a log-normal distribution
- Along the way: A gentle introduction to the statistical environment *R*

Claims Reserving Manual D5 (1997)

Faculty and Institute of Actuaries
Claims Reserving Manual v.2
(09/1997)
Section D5



[D5]
REGRESSION MODELS BASED ON LOG-INCREMENTAL PAYMENTS
Contributed by S Christofides

The first article in Volume 2 of this Manual by B Zehnwirth has shown the close connection between the intuitive Chain Ladder technique and the more formal two way analysis of variance model based on the log-incremental payments.

Models initiated by this more formal definition of the basic chain ladder have recently started to gain acceptance in loss reserving work and a number of papers on the subject have now been published. These models differ from the traditional techniques by a more formal definition of both the model assumptions and the parameter estimation and testing. With the formal models statistical estimates of reserves, that is both mean estimates and the associated model standard errors, can be calculated. The basic chain ladder is deterministic and produces point estimates of reserves.

The purpose of this paper is to serve as a basic introduction to these methods for the practitioner. To facilitate this a PC spreadsheet package is used to show how run-off models of the log-incremental payments can be identified and fitted in practice using multiple regression.

Source:
http://www.actuaries.org.uk/research-and-resources/documents/claims-reserving-manual-vol2-section-d5-regression-models-based-lo-0

Log-incremental Model

- Assume incremental claims payments P_{ij} follow a log-normal distribution
- $$\log\left(P_{ij}\right) = Y_{ij} \sim \mathcal{N}\left(\mu_{ij}, \sigma^2\right)$$
  $$E[Y_{ij}] = \mu_{ij} = a_i + b_j$$
  ← Linear model on a log-scale
  - One predictor for each origin and development period
- Transformation to original scale:
  $$E[P_{ij}] = \exp(\mu_{ij} + \sigma^2/2)$$
  $$\mathrm{Var}[P_{ij}] = E[P_{ij}]^2 \left(\exp(\sigma^2) - 1\right)$$

Note, coefficient of variation independent of mean.

Example Cumulative Data

origin	0	1	2	3	4	5	6
0	3,511	6,726	8,992	10,704	11,763	12,350	12,690
1	4,001	7,703	9,981	11,161	12,117	12,746	NA
2	4,355	8,287	10,233	11,755	12,993	NA	NA
3	4,295	7,750	9,773	11,093	NA	NA	NA
4	4,150	7,897	10,217	NA	NA	NA	NA
5	5,102	9,650	NA	NA	NA	NA	NA
6	6,283	NA	NA	NA	NA	NA	NA

Source: Claims Reserving Manual D5 (1997), page D5.16

Example Incremental Data

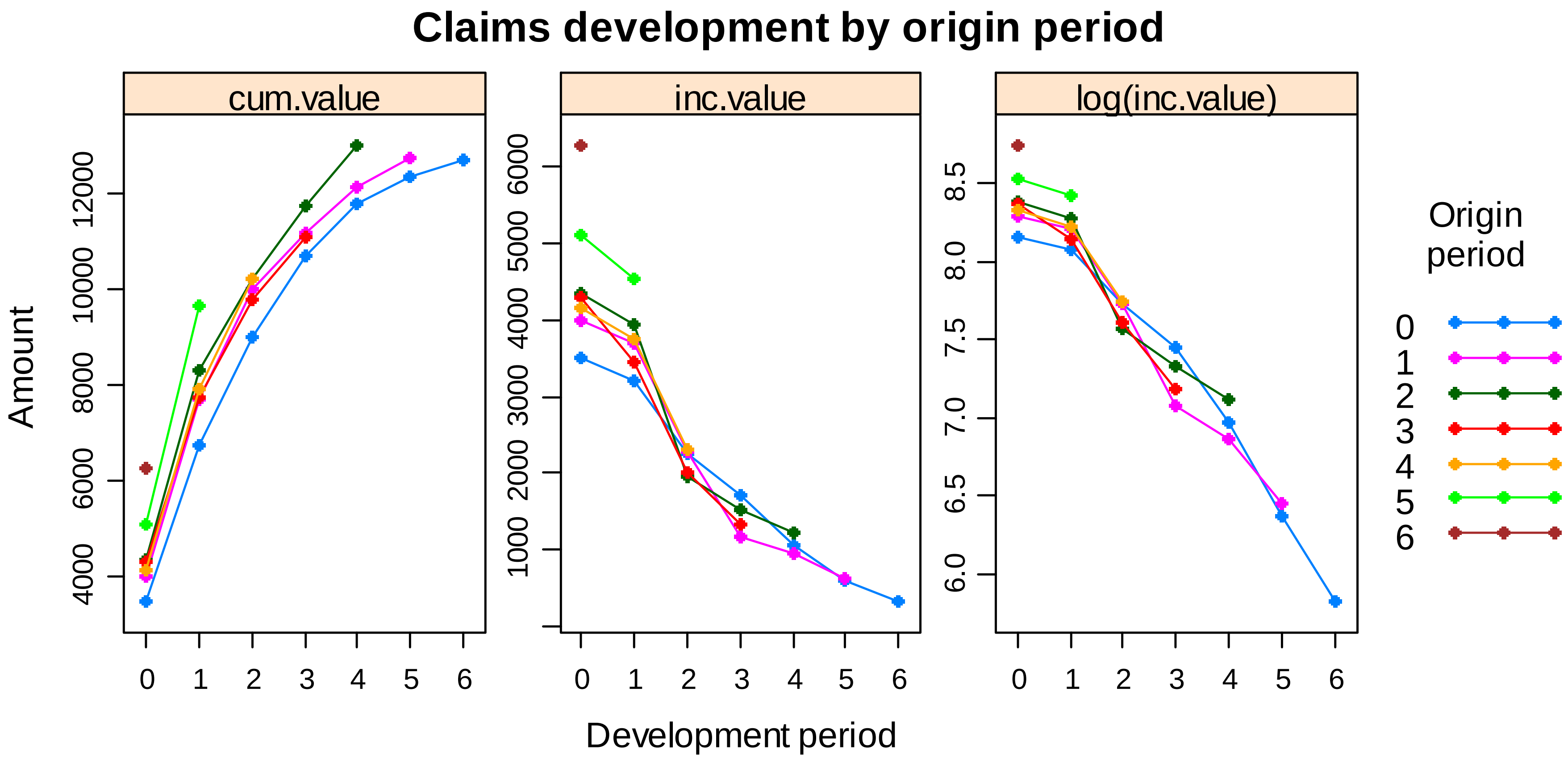
origin	0	1	2	3	4	5	6
0	3,511	3,215	2,266	1,712	1,059	587	340
1	4,001	3,702	2,278	1,180	956	629	NA
2	4,355	3,932	1,946	1,522	1,238	NA	NA
3	4,295	3,455	2,023	1,320	NA	NA	NA
4	4,150	3,747	2,320	NA	NA	NA	NA
5	5,102	4,548	NA	NA	NA	NA	NA
6	6,283	NA	NA	NA	NA	NA	NA

Source: Claims Reserving Manual D5 (1997), page D5.17

Insert Incremental Data into R

```
library(data.table)
dat <- data.table(
  originf=factor(rep(0:6, each=7)),
  devf=factor(rep(0:6, 7)),
  inc.value= c(3511,  3215,  2266,  1712, 1059, 587, 340,
               4001,  3702,  2278, 1180, 956, 629,  NA,
               4355,  3932,  1946, 1522, 1238,  NA,  NA,
               4295,  3455,  2023, 1320,   NA,  NA,  NA,
               4150,  3747,  2320,   NA,   NA,  NA,  NA,
               5102,  4548,    NA,   NA,   NA,  NA,  NA,
               6283,    NA,    NA,   NA,   NA,  NA,  NA))
```

Claims Development by Origin Period



Prepare Data for Modelling

- Initial model assumes
  - One predictor for each origin and development period
- Store years as categorical variables
  - Called ‘factors’ in R

Apply Linear Model on Log-transformed Incremental Claims

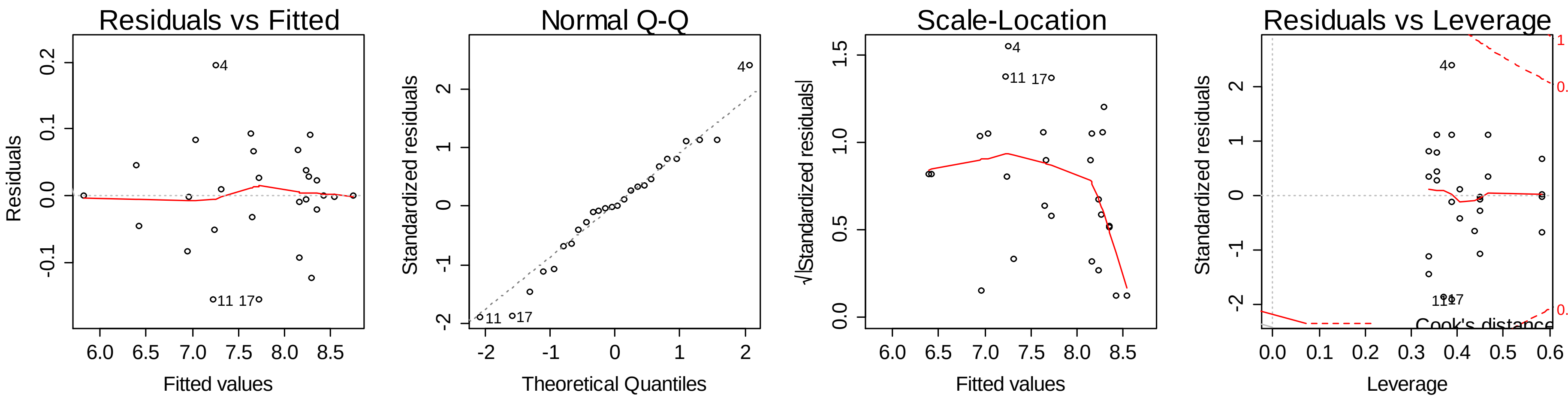
```
model <- lm(log(inc.value) ~
                  originf + devf,
              data=na.omit(dat))

library(arm)
display(model) #output
summary(model)$sigma
> [1] 0.1040994
```

Predictor	Coefficient	Coef.S.E.
Intercept	8.29	0.06
originf1	-0.02	0.06
originf2	0.07	0.06
originf3	-0.01	0.07
originf4	0.06	0.08
originf5	0.25	0.09
originf6	0.46	0.12
devf1	-0.12	0.06
devf2	-0.63	0.06
devf3	-1.04	0.07
devf4	-1.32	0.08
devf5	-1.87	0.09
devf6	-2.46	0.12

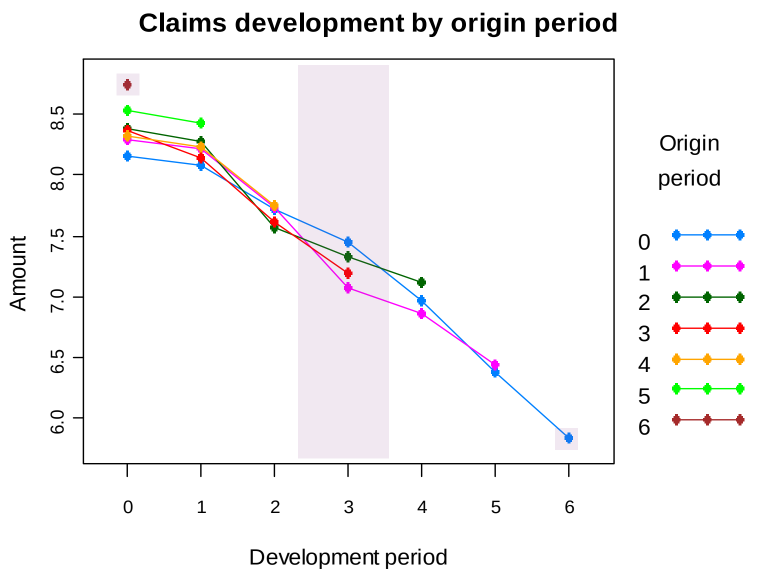
Review Residual Plots

```
par(mfrow=c(1,4))
plot(model)
```



Rows: 4, 7, 11, 28 are highlighted as potential outliers

Review data again



origin	0	1	2	3	4	5	6
0	3,511	3,215	2,266	**1,712**	1,059	587	**340**
1	4,001	3,702	2,278	**1,180**	956	629	NA
2	4,355	3,932	1,946	1,522	1,238	NA	NA
3	4,295	3,455	2,023	1,320	NA	NA	NA
4	4,150	3,747	2,320	NA	NA	NA	NA
5	5,102	4,548	NA	NA	NA	NA	NA
6	**6,283**	NA	NA	NA	NA	NA	NA

Row	Origin	Dev	Incr. Value
1	0	0	3511
2	0	1	3215
3	0	2	2266
4	**0**	**3**	**1712**
5	0	4	1059
6	0	5	587
7	**0**	**6**	**340**
8	1	0	4001
9	1	1	3702
10	1	2	2278
11	**1**	**3**	**1180**
12	1	4	956
13	1	5	629
14	2	0	4355
15	2	1	3932
16	2	2	1946
17	2	3	1522
18	2	4	1238
19	3	0	4295
20	3	1	3455
21	3	2	2023
22	3	3	1320
23	4	0	4150
24	4	1	3747
25	4	2	2320
26	5	0	5102
27	5	1	4548
28	**6**	**0**	**6283**

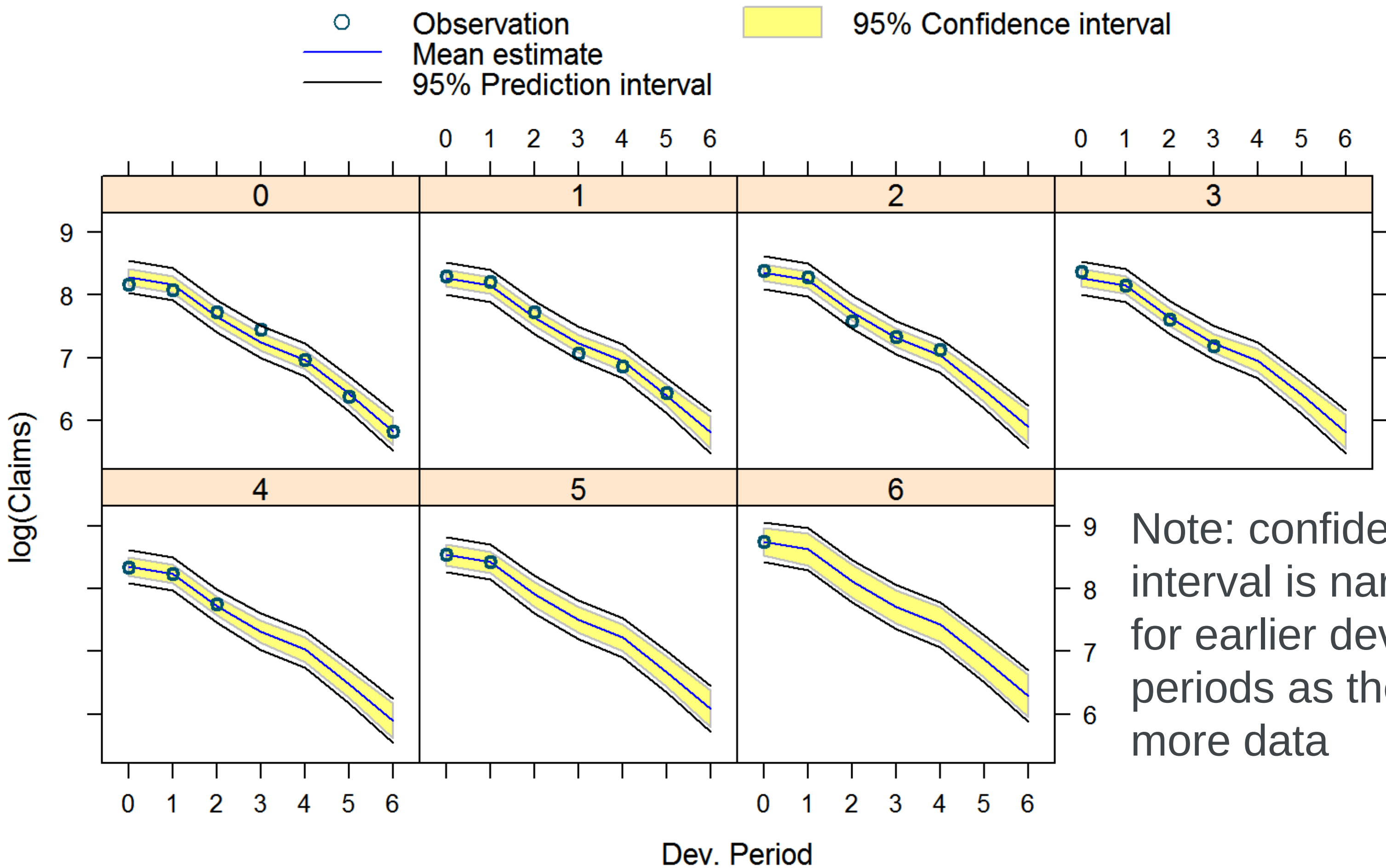
Mean versus Individual Prediction

- Individual prediction yields larger variance than mean prediction, and hence more uncertainty, wider confidence intervals, etc.
- In R, predict method for objects of class `lm` calls `predict.lm`; i.e., see `help(predict.lm)`, and the options for the interval argument:

mean prediction ("narrow")	interval="confidence"
individual prediction ("wide")	interval="prediction"

Source: Simon Jackman
http://jackman.stanford.edu/classes/350B/07/predictionforWeb.pdf

Review Prediction and Confidence Intervals by Origin Period (log-scale)



Note: confidence interval is narrower for earlier dev. periods as there is more data

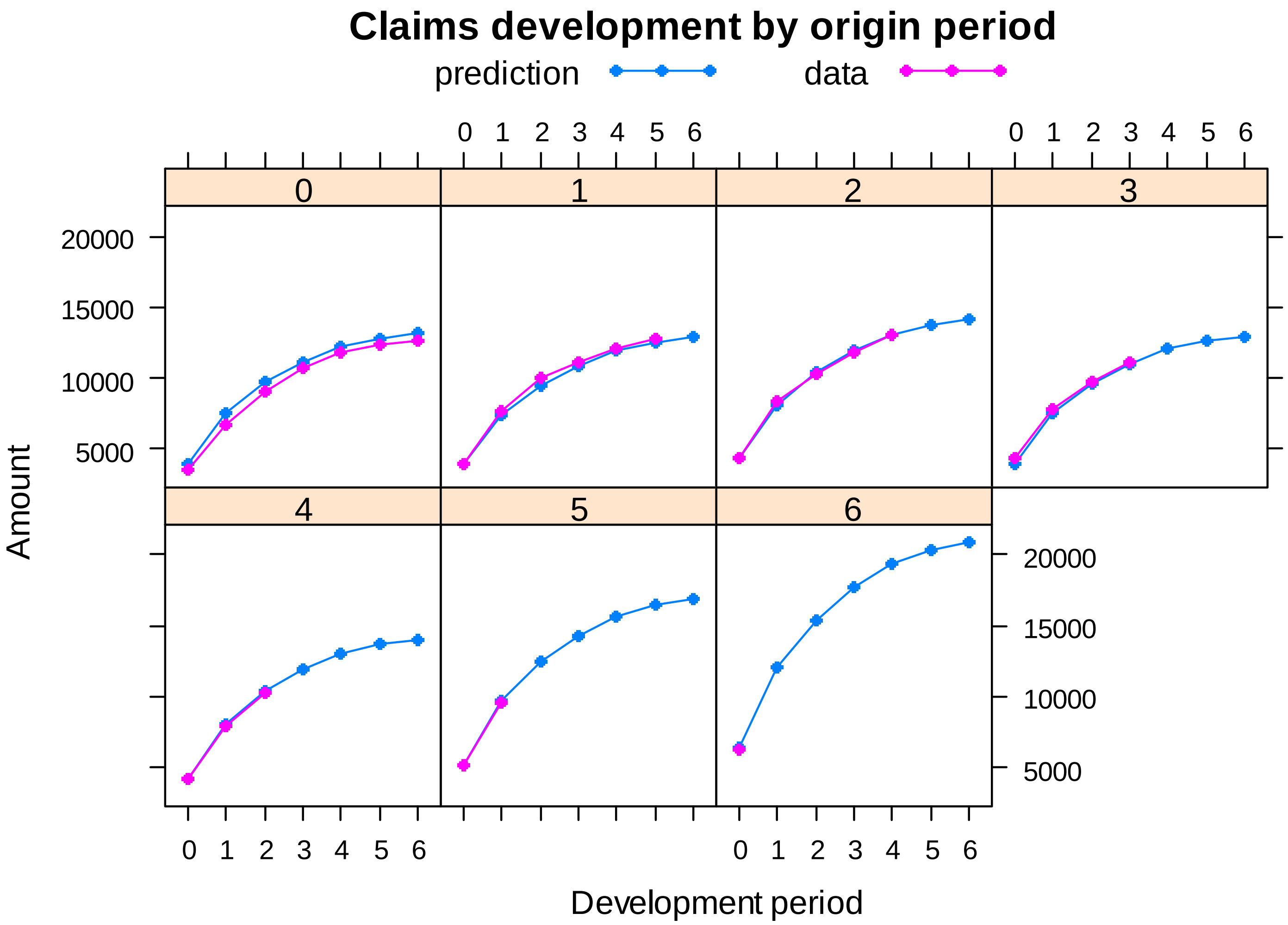
Out-of Sample Prediction

- For predictions of data on the ***original scale*** consider:
  - Residual variance $\hat{\sigma}^2$ (Estimated variance of the random/stochastic error aka 'process error')
  - Prediction error $s.e.(\hat{Y}_{ij})$ (aka 'parameter error')

- Hence: $\hat{P}_{ij} = \exp\left(\hat{Y}_{ij} + \frac{1}{2}(\hat{\sigma}^2 + s.e.(\hat{Y}_{ij})^2)\right)$

```
pred.dat <- dat[, c('originf', 'devf', 'inc.value')]
pred <- predict(model, newdata=pred.dat, se.fit=TRUE)
phat <- exp(pred$fit + 0.5*(pred$residual.scale^2 + pred$se.fit^2))
sum(phat[is.na(pred.dat$inc.value)])
[1] 28823.4
```

Compare Prediction with Data



Prediction Error on Original Scale

- Recall: $\mathrm{Var}[P_{ij}] = E[P_{ij}]^2 \left(\exp(\sigma^2) - 1\right)$

- Hence: $s.e.(\hat{P}_{ij}) = \hat{P}_{ij}\sqrt{\exp(\hat{\sigma}^2 + s.e.(\hat{Y}_{ij})^2) - 1}$

```
s.e.phat <- phat * sqrt(exp(pred$residual.scale^2 +
                                        pred$se.fit^2)-1)
```

Prediction Error of Total Reserves

$$D := \text{Future desgin matrix}$$
$$\Omega := \text{Covariance matrix of the errors}$$

$$s.e.(\sum_{i,j \in \text{future}} \hat{P}_{ij}) = \sqrt{\hat{P}_{ij}^T \cdot \exp(D \cdot \Omega \cdot D^T - I) \cdot \hat{P}_{ij}}$$

```
future.dev <- which(is.na(dat$inc.value))
fdm <- model.matrix(~ originf + devf, data=pred.dat[future.dev,])
varcovar <- fdm %*% vcov(model) %*% t(fdm)
(Total.SE <- sqrt(t(phat[future.dev]) %*% (exp(varcovar)-1) %*%
                                              phat[future.dev]))
```

> 1971.238

Prediction Summary

- Overall predicted reserve: 28,823
- Overall s.e.(Reserve): 1,971

Tweedie GLM for Comparison

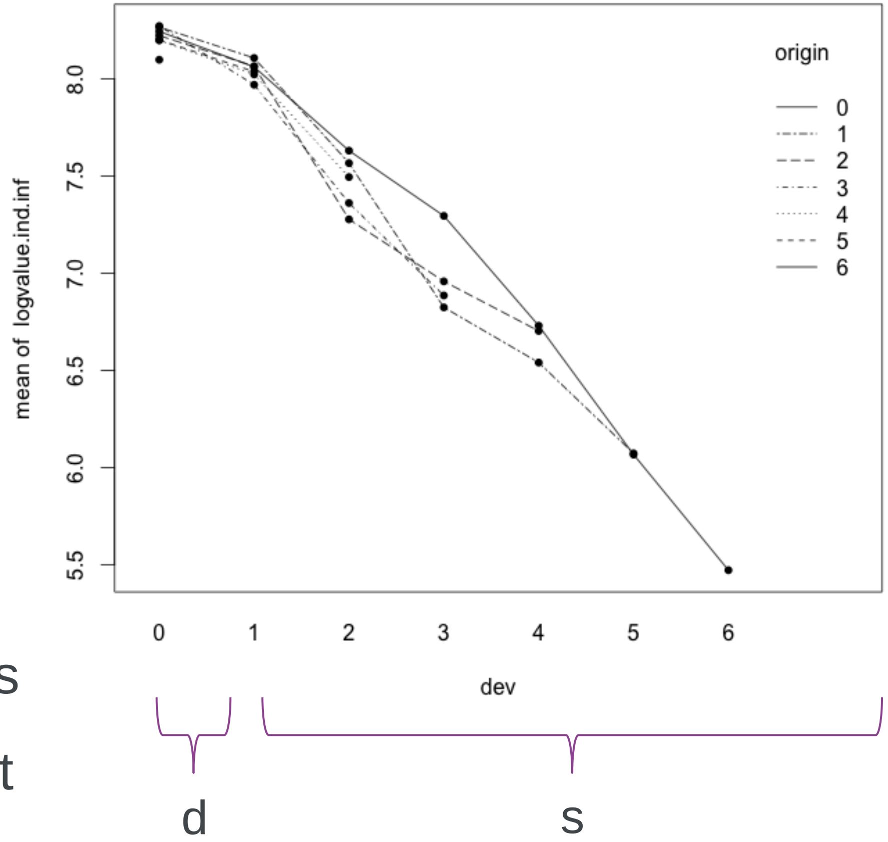
```
library(ChainLadder)
glmReserve(UKMotor, var.power = NULL)
      Latest Dev.To.Date Ultimate  IBNR       S.E         CV
2008   12746   0.9732000    13097   351  110.0539 0.31354398
2009   12993   0.9260870    14030  1037  176.9362 0.17062310
2010   11093   0.8444089    13137  2044  238.5318 0.11669853
2011   10217   0.7360951    13880  3663  335.6824 0.09164140
2012    9650   0.5739948    16812  7162  543.6473 0.07590719
2013    6283   0.3038201    20680 14397 1098.7989 0.07632138
total  62982   0.6873063    91636 28654 1622.4617 0.056622
```

Mack's Model for Comparison

```
library(ChainLadder)
MackChainLadder(Triangle = UKMotor, est.sigma = "Mack")
     Latest Dev.To.Date Ultimate   IBNR Mack.S.E CV(IBNR)
2007 12,690       1.000   12,690      0     0.00      NaN
2008 12,746       0.973   13,097    351     3.62   0.0103
2009 12,993       0.926   14,031  1,038    22.90   0.0221
2010 11,093       0.844   13,138  2,045   141.98   0.0694
2011 10,217       0.736   13,880  3,663   426.70   0.1165
2012  9,650       0.574   16,812  7,162   692.39   0.0967
2013  6,283       0.304   20,680 14,397   900.58   0.0626
              Totals
Latest:    75,672.00
Dev:            0.73
Ultimate: 104,327.77
IBNR:      28,655.77
Mack.S.E    1,417.27
CV(IBNR):       0.05
```

Steps to Improve the Log-transformed Model

- Index triangle for changes in
  - Exposure
  - Earnings
  - Inflation
- Test for reduction in number of variables, e.g.
  - No origin period trends
  - Constant development trend after year 1

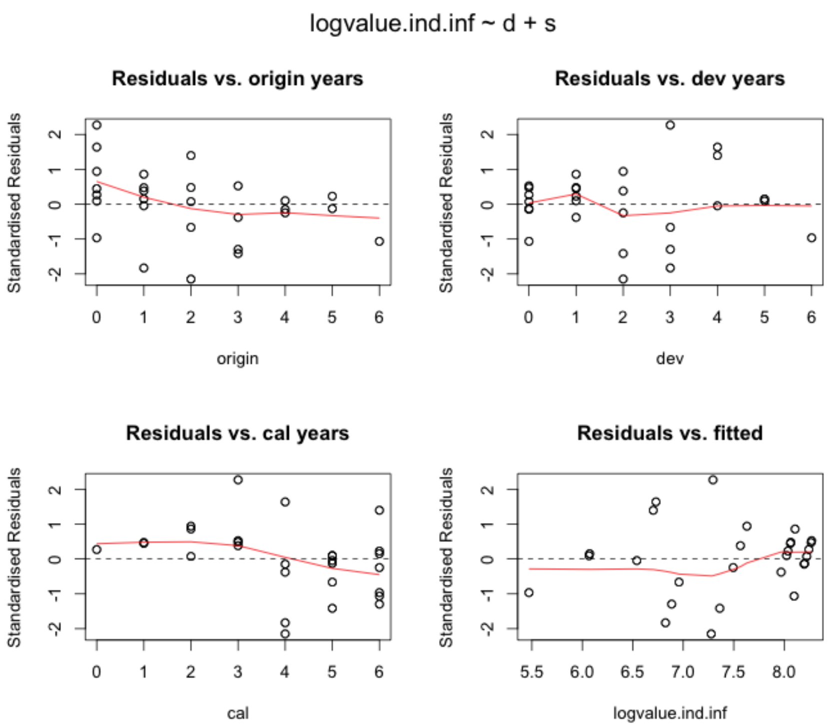


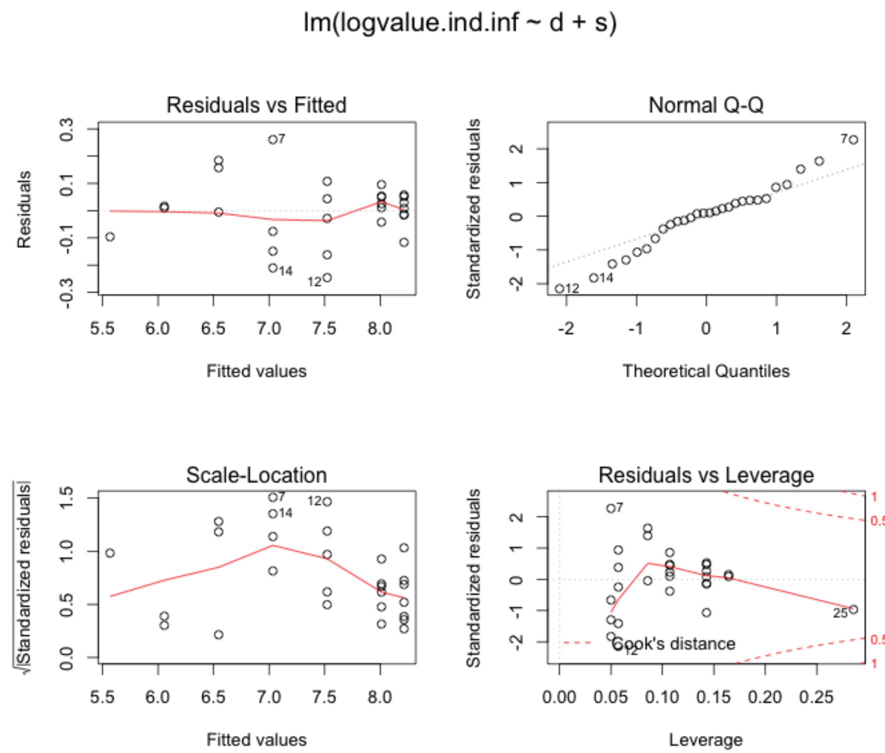
Updated Model

```
# Page D5.39
display(model2 <- lm(logvalue.ind.inf ~  d + s, data=dat))
lm(formula = logvalue.ind.inf ~ d + s, data = na.omit(dat))
            coef.est coef.se
(Intercept)  8.50     0.05
d           -0.29     0.07
s           -0.42     0.02
---
n = 28, k = 3
residual sd = 0.12, R-Squared = 0.97
```

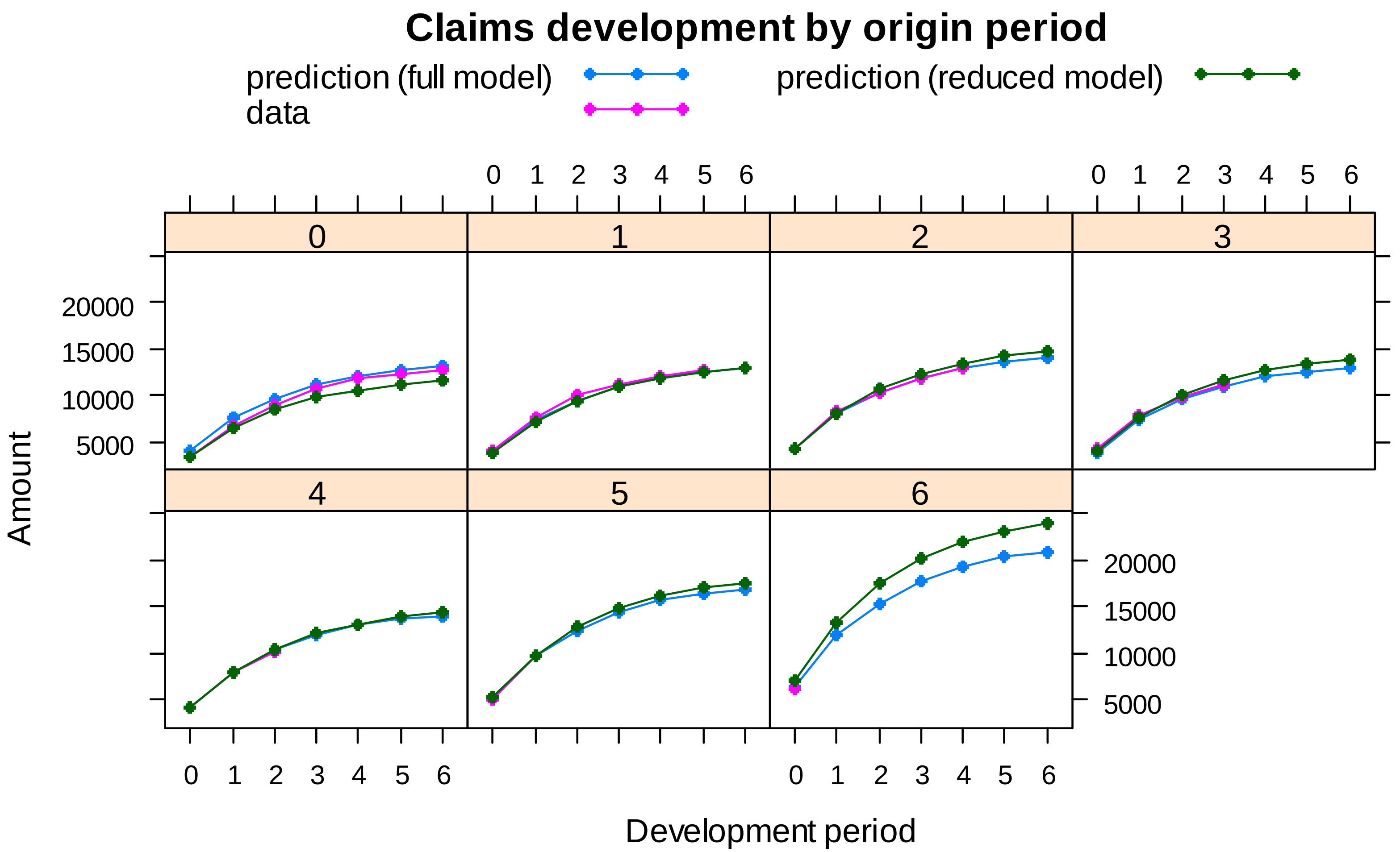
Example taken from http://www.magesblog.com/2013/01/reserving-based-on-log-incremental_22.html

Review Residual Plots





Compare Predictions with Data



Model Prediction with an Additional Tail of 6 Years

```
FM4 <- log.incr.predict(Fit4, ND,
                        claims.inflation=0.075,
                        volume.index="volume.index")
# Page D5.41
FM4$Totals
#  Total.Reserve Total.SE         CV
#       38083.25 1724.987 0.04529515
```

Data frame with future development years. No change in model.

Example taken from http://www.magesblog.com/2013/01/reserving-based-on-log-incremental_22.html

Conclusions

- Working with triangles in a ‘long format’ makes statistical modelling often ‘easier’
- Using a statistical tool facilitates:
  - Quick testing for different models, e.g.
    - Log-transform, Poisson, Tweedie, Mixed–effects, etc.
  - Forecasting models
  - Estimating prediction errors
- Remember: Science aims to prove models wrong

Additional Resources

- Claims Reserving Manual D5 (1997)
  - Series of blog posts to reproduce results in R:
    - http://www.magesblog.com/2013/01/reserving-based-on-log-incremental.html
- A Practitioner's Introduction to Stochastic Reserving:
    - http://mages.github.io/PSRWP
- ChainLadder: R package for stochastic reserving
    - https://cran.r-project.org/package=ChainLadder
- Computational Actuarial Science with R
    - edited by Arthur Charpentier

Questions

Comments

The views expressed in this presentation are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation.