2019 Predictive Analytics Symposium

Session 30: M/S - General Insurance Applications of Predictive Analytics (PA)

Mixed Effects Models

Brian Hartman, ASA, PhD, Brigham Young University
Peng Shi, FSA, PhD, University of Wisconsin-Madison

Outline

1. Longitudinal Data
2. Linear Mixed Effects Models
3. Generalized Linear Mixed Effects Models

Outline

What are longitudinal and panel data?

- Regression is a statistical tool to study the distribution of an outcome of interest in terms of other variables
- In regression, measurements are recorded at the level of research unit or observational unit
  - Unit of analysis is referred to as individuals in econometrics literature and subjects in statistics literature
  - Dependent variable: outcome of interest
  - Explanatory variable: other measurements
- In insurance and actuarial applications, one could think of a policyholder as an subject.

What are longitudinal and panel data?

- Model building often depends on the type of data
  - Cross-sectional data focuses on static relationship, while Time series data examines the dynamic relationship.
  - Longitudinal/panel data: a marriage of cross-sectional and time series data
    - measurements on a cross-section of subjects
    - repeated observed over time (multiple observations on each subject)
  - We follow each of n subjects for a maximum of $T = \max\{T_1, \ldots, T_n\}$ time periods
    - 1st subject $\{y_{11}, y_{12}, \ldots, y_{1T_1}\}$
    - 2nd subject $\{y_{21}, y_{22}, \ldots, y_{1T_2}\}$
    - $\ldots$
    - nth subject $\{y_{n1}, y_{n2}, \ldots, y_{nT_n}\}$

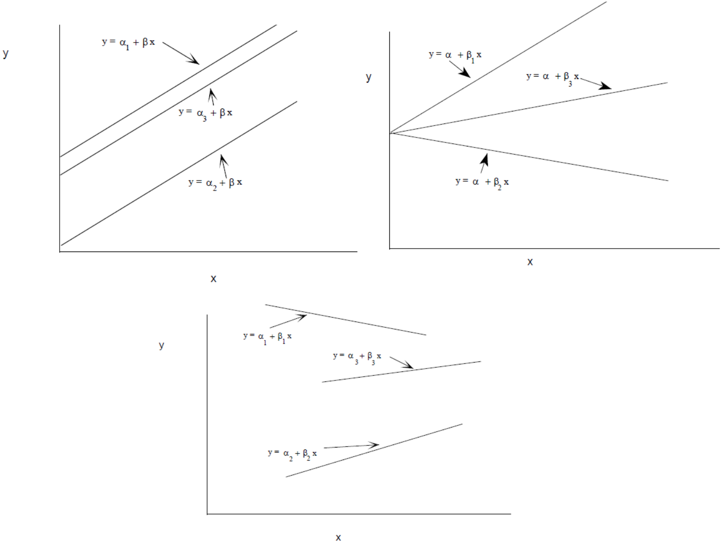
What are longitudinal and panel data?

- Consider some applications that actuaries might face that fall into this framework
  - Personal lines insurance: y_{it} is the number of claims of a policyholder.
  - Commercial lines insurance: y_{it} is the loss ratio of a customer.
  - Insurance sales: y_{it} is the sales of an insurance agent.
  - Coverage selection: y_{it} is the coverage selected by the policyholder, e.g. deductible level.
  - Customer Retention: y_{it} is the indicator whether a policyholder purchases coverage.

Benefits of longitudinal data

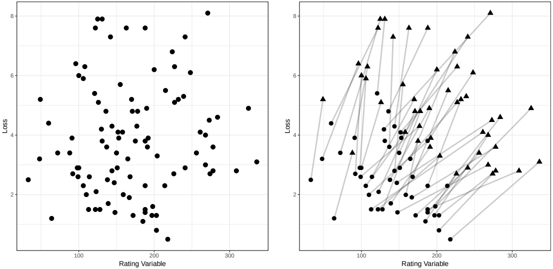
- Control for individual heterogeneity
  - Heterogeneity means that subjects are different and unique
  - Panel data allows us to account for the uniqueness through subject-specific parameters, such as
    - heterogeneous intercept: $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$
    - heterogeneous slope: $y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}$
    - both intercept and slope: $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}$

Benefits of longitudinal data



Benefits of longitudinal data

- Study dynamic relationship
  - In cross-sectional regression, actuaries makes inference about effect of explanatory variables on the dependence variable. It is a static relationship because there is no time element in the inference.
  - In longitudinal regression, actuaries are interested in changes over time, known as dynamic relationship.



Benefits of longitudinal data

- Efficiency and information sharing
  - Actuaries would be able to obtain more precise (efficient) estimates of parameters. Suppose observations from different years are independent, a large sample means:
    - more variability in the data and thus less collinearity among variables
    - more degrees of freedom and thus less uncertainty in estimates
  - Longitudinal data often exhibit features of clustering. By exploring and incorporating this relationship into model, actuaries would also be able to make more efficient prediction.

Benefits of longitudinal data

- Identify and measure effects that are usually not detectable in pure cross-section or pure time series data
  - Example: we are interested in the accident rate in car insurance
  - In cross-sectional data, we have $\mathrm{E}(y_i|\boldsymbol{x}_i) = \mathrm{E}(y_j|\boldsymbol{x}_j) = 0.5$. Two possibilities:
    - subjects i and j are from homogeneous population, both 50% chance of accident in a given year
    - subjects i and j are from heterogeneous population, i has accident every year and j doses not have accident at all
  - With longitudinal data, two groups can be separated by investigating the effect of lagged dependent variable

Outline

Linear Mixed Effects Model

- The linear mixed effects is

$$y_{it} = \boldsymbol{z}_{it}'\boldsymbol{\alpha}_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

- It contains both fixed effects and random effects
- This model allows for subject-specific intercept and slopes
- The model has matrix representation

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

Linear Mixed Effects Model

- Sampling assumptions
  - Conditional on $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$, $\{\boldsymbol{y}_i\}$ are independent random vectors
  - $\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{\alpha}_i) = \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{X}_i\beta$
  - $\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{\alpha}_i) = \boldsymbol{R}_i$
  - $\mathrm{E}(\boldsymbol{\alpha}_i) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\alpha}_i) = \boldsymbol{D}$, and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ are i.i.d.
- This implies:
  - $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{R}_i$. It allows for heteroscedasticity and serial correlation.
  - Subject-specific effects and the noise term are uncorrelated, i.e. $\mathrm{Cov}(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}_i') = 0$
- Additional assumptions for finite sample inference
  - Conditional on $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$, $\{\boldsymbol{y}_i\}$ are normally distributed
  - $\{\boldsymbol{\alpha}_i\}$ is normally distributed

Linear Mixed Effects Model

- The marginal model is:

$$\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$
$$\mathrm{Var}(\boldsymbol{y}_i) = \boldsymbol{V}_i(\boldsymbol{\tau}) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i$$

- The linear mixed model implies the above marginal model, but not vice versa.
- Marginal model is useful when the interest is the estimation of fixed effects, while linear mixed effect model should be used when the interest in the prediction.

Linear Mixed Effects Model

- Special Cases:
  - One-way ANOVA model

$$y_{it} = \mu + \alpha_i + \varepsilon_{it}$$

  - In general, error components model

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

- Key assumptions:
  - $\{\alpha_i\}$ are i.i.d with zero mean and variance σ_α^2
  - $\{\alpha_i\}$ and error $\{\varepsilon_{it}\}$ term are uncorrelated
  - homoscedasticity $\mathrm{Var}(\varepsilon_{it}) = \sigma^2$
  - no serial correlation $\mathrm{Cov}(\varepsilon_{ir}, \varepsilon_{is}) = 0$

Linear Mixed Effects Model

- Estimation
  - Fixed effects $\boldsymbol{\beta}$ are estimated using GLS
  - Variance components are estimated using MLE or REML
- Prediction. Suppose we wish to predict a random variable w, where $\mathrm{E}(w) = \boldsymbol{\lambda}'\boldsymbol{\beta}$ and $\mathrm{Var}(w) = \sigma_w^2$. Given $\boldsymbol{\beta}$, the best linear predictor of w is (in terms of MSE)

$$w^* = \boldsymbol{\lambda}'\boldsymbol{\beta} + \mathrm{Cov}(w, \boldsymbol{y})\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

  - the results do not not rely on distributional assumption
  - under normality, one can show $w^* = \mathrm{E}(w|\boldsymbol{y})$
- One could use `R package lme4` for implementation.

Linear Mixed-Effects Models

- Consider an application of linear mixed-effects model
- Recall the response is generated by $\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$
- We are interested in subject i in period $T_i + L$ (L lead time units in the future)
- Two quantities of interest
  - Conditional mean $\mathrm{E}(y_{i,T_i+L}|\boldsymbol{\alpha}_i) = \boldsymbol{z}'_{i,T_i+L}\boldsymbol{\alpha}_i + \boldsymbol{x}'_{i,T_i+L}\boldsymbol{\beta}$
  - Future response $y_{i,T_i+L} = \boldsymbol{z}'_{i,T_i+L}\boldsymbol{\alpha}_i + \boldsymbol{x}'_{i,T_i+L}\boldsymbol{\beta} + \varepsilon_{i,T_i+L}$

Linear Mixed-Effects Models

- Conditional mean. The BLUP is

$$w_{BLUP} = \boldsymbol{x}'_{i,T_i+L}\boldsymbol{b}_{GLS} + \boldsymbol{z}'_{i,T_i+L}\boldsymbol{a}_{i,BLUP}$$

- Future response. The BLUP is

$$w_{BLUP} = \boldsymbol{x}'_{i,T_i+L}\boldsymbol{b}_{GLS} + \boldsymbol{z}'_{i,T_i+L}\boldsymbol{a}_{i,BLUP} + \mathrm{Cov}(\varepsilon_{i,T_i+L}, \boldsymbol{\varepsilon}_i)\boldsymbol{R}_i^{-1}\boldsymbol{e}_{i,BLUP}$$

  - Special case when $\boldsymbol{R}_{i,st} = \sigma^2\rho^{|s-t|}$

$$w_{BLUP} = \boldsymbol{x}'_{i,T_i+L}\boldsymbol{b}_{GLS} + \boldsymbol{z}'_{i,T_i+L}\boldsymbol{a}_{i,BLUP} + \rho^L e_{iT,BLUP}$$

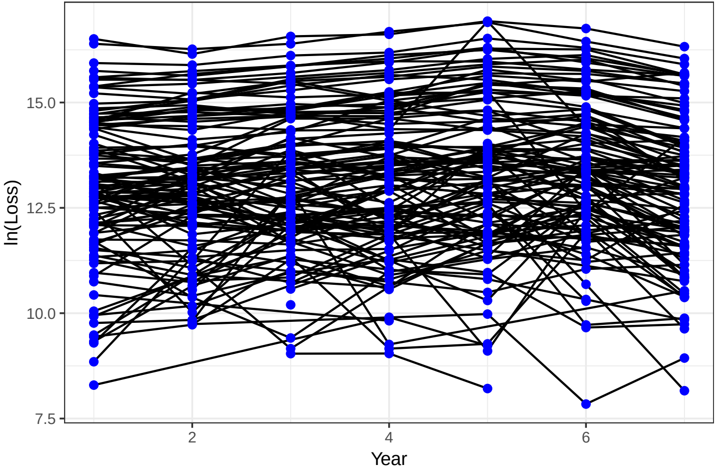
  - Further if $\rho = 0$, the point prediction is the same as the case of conditional mean.

Linear Mixed-Effects Models

- We consider the loss data in Worker's Compensation Insurance
  - The data are from the National Council on Compensation Insurance
  - It contains losses due to permanent partial disability (see Klugman (1992))
  - 118 occupation or risk classes are observed over 7 years
- The variable of interest is `Loss`. Possible explanatory variables are `Year` and `Payroll`
  - We use `Payroll` as an offset.

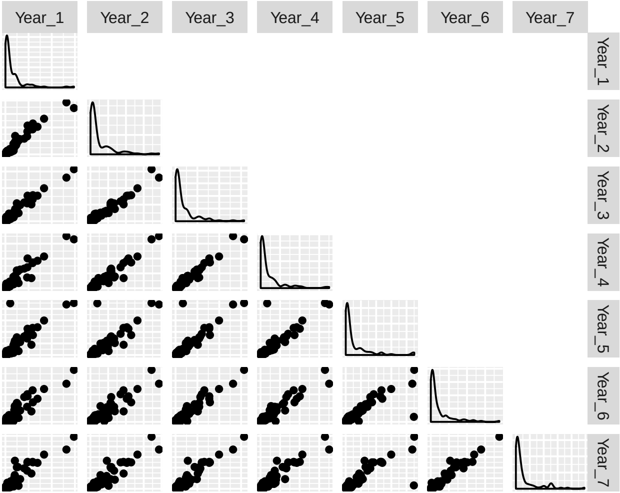
Linear Mixed-Effects Models

- The time series plot of `Loss`:



Linear Mixed-Effects Models

- The scatter plot of Loss:



Linear Mixed-Effects Models

- We consider the following modeling strategies:
  - Pooled regression:

$$\log(Loss_{it}) = \log(Payroll_{it}) + \beta_0 + \beta_1 Year_{it} + \varepsilon_{it}$$

  - Fixed effect model:

$$\log(Loss_{it}) = \log(Payroll_{it}) + \beta_{0,i} + \beta_1 Year_{it} + \varepsilon_{it}$$

  - Error-component model:

$$\log(Loss_{it}) = \log(Payroll_{it}) + \alpha_i + \beta_0 + \beta_1 Year_{it} + \varepsilon_{it}$$

  - Random coefficient model:

$$\log(Loss_{it}) = \log(Payroll_{it}) + \alpha_{0,i} + \alpha_{1,i} + \beta_0 + \beta_1 Year_{it} + \varepsilon_{it}$$

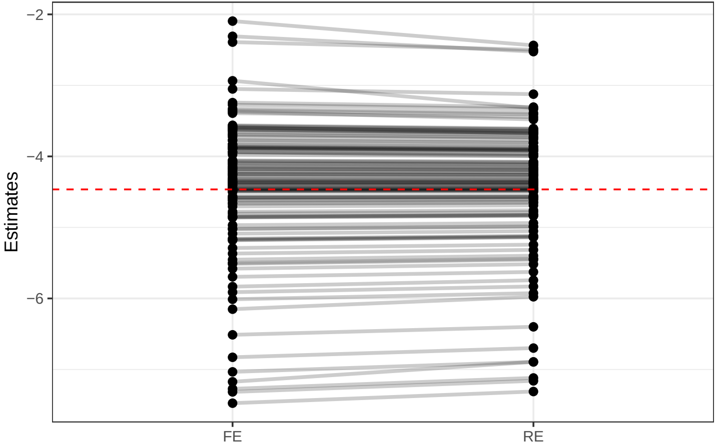
Linear Mixed-Effects Models

- Goodness-of-fit statistics are:

Model	df	LogLik	AIC	BIC
Pooled CS	3	-988.35	1982.69	1996.20
Fixed Effects	120	-482.00	1204.00	1744.69
Error Component	4	-720.23	1448.46	1466.48
Random Coefficient	6	-719.35	1450.70	1477.74

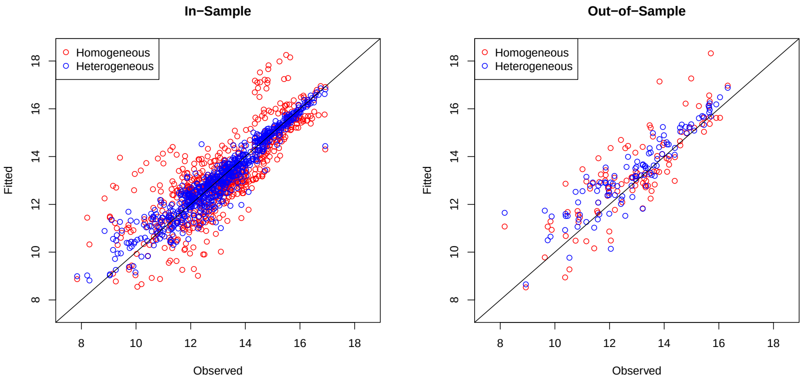
Linear Mixed-Effects Models

- Comparison of different estimators:



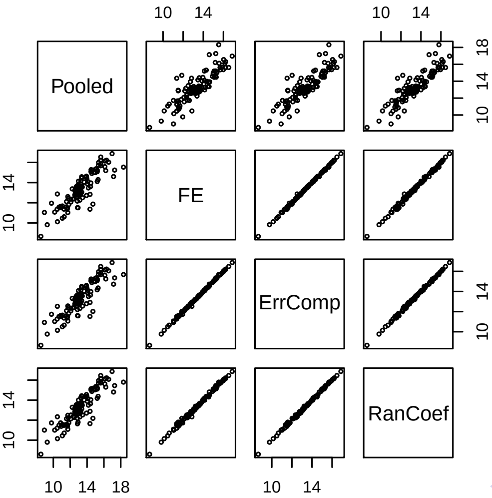
Linear Mixed-Effects Models

- Comparison of homogeneous and heterogeneous models:



Linear Mixed-Effects Models

- Comparison of out-of-sample prediction:



Linear Mixed-Effects Models

- Hold-out sample validation statistics are:

	Pooled	FE	ErrComp	RanCoeff
MAE†	2.137	0.704	0.687	0.691
RMSE†	9.093	1.463	1.433	1.420
MAPE	0.685	0.517	0.507	0.498
Pearson	0.505	0.968	0.970	0.974
Spearman	0.823	0.917	0.922	0.927

† in millions of dollars.

Linear Mixed-Effects Models

- Linear mixed effects models enhance the application of credibility theory by incorporating covariates.
- Several well-known credibility models can be viewed in the framework of linear mixed effects models.

$$y_{it} = \boldsymbol{z}'_{it}\boldsymbol{\alpha}_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

  - Bühlmann: $\boldsymbol{z}_{it} = \boldsymbol{x}_{it} = 1$, $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{I}_{T_i}$
  - Bühlmann-Straub: $\boldsymbol{z}_{it} = \boldsymbol{x}_{it} = (1, t)'$, $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2 diag(1/w_{i1}, \ldots, 1/w_{iT_i})$
  - Hachemeister: $\boldsymbol{z}_{it} = \boldsymbol{x}_{it} = (1, t)'$, $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{I}_{T_i}$

Outline

Generalized Linear Mixed-Effects Models

- This section extends the mixed-effects models to outcomes with a distribution from the exponential family.
- *Definition.* The distribution of the *linear exponential family* is

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + S(y,\phi)\right).$$

  - y is a dependent variable and θ is the parameter of interest.
  - ϕ is a scale parameter that we often will assume is known.
  - $b(\theta)$ depends only on the parameter θ, not the dependent variable.
  - $S(y,\phi)$ is a function of the dependent variable and the scale parameter, not the parameter θ.
- We can show that

$$\mu = \mathrm{E}\ y = b'(\boldsymbol{\theta}) \quad \text{and} \quad \mathrm{Var}\ y = \phi b''(\boldsymbol{\theta}) = \phi V(\mu),$$

where $V(\cdot)$ is known as variance function.

Linear Exponential Family of Distributions

Table: Selected Distributions of the One-Parameter Exponential Family

Distribution	Parameters	Density or Mass Function	Components	E y	Var y
General	$\theta,\ \phi$	$\exp\left(\frac{y\theta - b(\theta)}{\phi} + S(y,\phi)\right)$	$\theta,\ \phi, b(\theta), S(y,\phi)$	$b'(\theta)$	$b''(\theta)\phi$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$	$\mu, \sigma^2, \frac{\theta^2}{2}, -\left(\frac{y^2}{2\phi} + \frac{\ln(2\pi\phi)}{2}\right)$	$\theta = \mu$	$\phi = \sigma^2$
Binomal	π	$\binom{n}{y}\pi^y(1-\pi)^{n-y}$	$\ln\left(\frac{\pi}{1-\pi}\right)$, 1, $n\ln(1+e^\theta)$, $\ln\binom{n}{y}$	$n\frac{e^\theta}{1+e^\theta}$ $= n\pi$	$n\frac{e^\theta}{(1+e^\theta)^2}$ $= n\pi(1-\pi)$
Poisson	λ	$\frac{\lambda^y}{y!}\exp(-\lambda)$	$\ln\lambda, 1, e^\theta, -\ln(y!)$	$e^\theta = \lambda$	$e^\theta = \lambda$
Gamma	α, β	$\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}\exp(-y\beta)$	$-\frac{\beta}{\alpha}, \frac{1}{\alpha}, -\ln(-\theta), -\phi^{-1}\ln\phi$ $-\ln\left(\Gamma(\phi^{-1})\right) + (\phi^{-1}-1)\ln y$	$-\frac{1}{\theta} = \frac{\alpha}{\beta}$	$\frac{\phi}{\theta^2} = \frac{\alpha}{\beta^2}$
Inverse Gaussian	μ, λ	$\sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right)$	$-\mu^2/2, 1/\lambda, -\sqrt{-2\theta}$, $\theta/(\phi y) - 0.5\ln(\phi 2\pi y^3)$	$(-2\theta)^{-1/2}$ $= \mu$	$\phi(-2\theta)^{-3/2}$ $= \frac{\mu^3}{\lambda}$

Variance as a Function of the Mean

Table: Variance Functions for Selected Distributions

Distribution	Variance Function $v(\mu)$
Normal	1
Bernoulli	$\mu(1-\mu)$
Poisson	μ
Gamma	μ^2
Inverse Gaussian	μ^3

- The choice of the variance function drives many inference properties, not the choice of the distribution.

Generalized Linear Mixed-Effects Models

- The generalized linear mixed-effects model is specified as:

$$
\begin{aligned}
y_{it}|\boldsymbol{\alpha}_i &\sim f(y_{it}|\boldsymbol{\alpha}_i), \\
f(y_{it}|\boldsymbol{\alpha}_i) &= \exp\left(\frac{y_{it}\theta_{it} - b(\theta_{it})}{\phi} + S\left(y_{it}, \phi\right)\right). \\
\boldsymbol{\alpha}_i &\sim p(\boldsymbol{\alpha}_i)
\end{aligned}
$$

- The conditional mean and variance are:

$$\mu_{it} = \mathrm{E}(y_{it}|\boldsymbol{\alpha}_i) = b'(\boldsymbol{\theta}_{it}) \quad \text{and} \quad \mathrm{Var}(y_{it}|\boldsymbol{\alpha}_i) = \phi b''(\boldsymbol{\theta}_{it}) = \phi V(\mu_{it})$$

- Both fixed and random effects are specified via link function:

$$g(\mu_{it}) = \boldsymbol{z}_{it}'\boldsymbol{\alpha}_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}$$

Generalized Linear Mixed-Effects Models

- The marginal mean and variance are:

$$\begin{aligned} \mathrm{E}(y_{it}) &= \mathrm{E}(\mathrm{E}(y_{it}|\boldsymbol{\alpha}_i)) = \mathrm{E}(g^{-1}(\mathbf{z}_{it}'\boldsymbol{\alpha}_i + \mathbf{x}_{it}'\boldsymbol{\beta})) \\ \mathrm{Var}(y_{it}) &= \mathrm{Var}(\mathrm{E}(y_{it}|\boldsymbol{\alpha}_i)) + \mathrm{E}(\mathrm{Var}(y_{it}|\boldsymbol{\alpha}_i)) \\ &= \mathrm{Var}(g^{-1}(\mathbf{z}_{it}'\boldsymbol{\alpha}_i + \mathbf{x}_{it}'\boldsymbol{\beta})) + \mathrm{E}(\phi V(g^{-1}(\mathbf{z}_{it}'\boldsymbol{\alpha}_i + \mathbf{x}_{it}'\boldsymbol{\beta}))) \end{aligned}$$

- The regression coefficients do not have a marginal interpretation, it measures the effects conditional on the random effects.

Generalized Linear Mixed-Effects Models

- Estimation. The fixed effects parameters can be estimated using MLE:

$$l(\boldsymbol{\beta}, \boldsymbol{\tau}) = \prod_{i=1}^{n} \int \prod_{t=1}^{T_i} f(y_{it}|\boldsymbol{\alpha}_i) p(\boldsymbol{\alpha}_i) d\boldsymbol{\alpha}.$$

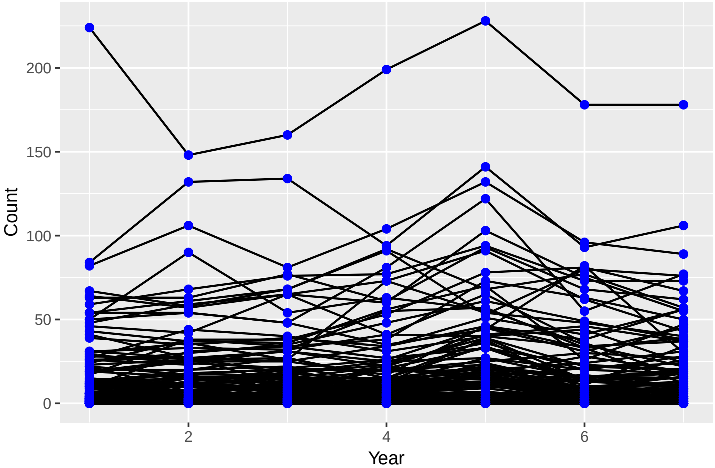
  - This requires approximations in the estimation.
  - Only conjugate distributions lead to closed-form solution.
- Prediction. The inference of random effects $\boldsymbol{\alpha}_i$ is via empirical Bayes.
- Alternatively, one could perform a full Bayesian approach to the mixed effects models.
- One could use `R package lme4` for implementation.

Generalized Linear Mixed-Effects Models

- We consider the claim frequency in Worker's Compensation Insurance
  - The data are from the National Council on Compensation Insurance
  - Claim frequency are observed on a yearly basis (see Klugman (1992))
  - 130 occupation or risk classes are observed over 7 years
- The variable of interest is `Count`. Possible explanatory variables are `Year` and `Payroll`
  - We use `Payroll` as an offset.

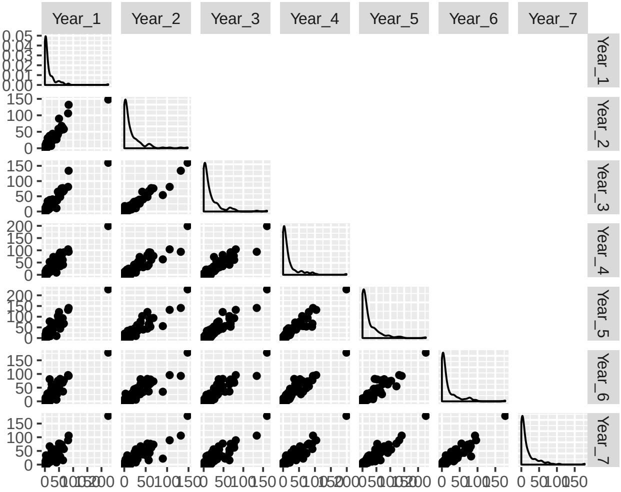
Generalized Linear Mixed-Effects Models

- The time series plot of `Count`:



Generalized Linear Mixed-Effects Models

- The scatter plot of `Count`:



Generalized Linear Mixed-Effects Models

- We consider the Poisson regression model with a log link function:

$$Count_{it} \sim Poisson(\lambda_{it})$$

  - Pooled regression:

$$\lambda_{it} = Payroll_{it} \exp(\beta_0 + \beta_1 Year_{it})$$

  - Fixed effects model:

$$\lambda_{it} = Payroll_{it} \exp(\beta_{0,i} + \beta_1 Year_{it})$$

  - Random intercept model:

$$\lambda_{it} = Payroll_{it} \exp(\alpha_i + \beta_0 + \beta_1 Year_{it})$$

  - Random intercept/slope model:

$$\lambda_{it} = Payroll_{it} \exp(\alpha_{0,i} + \alpha_{1,i} Year_{it} + \beta_0 + \beta_1 Year_{it})$$

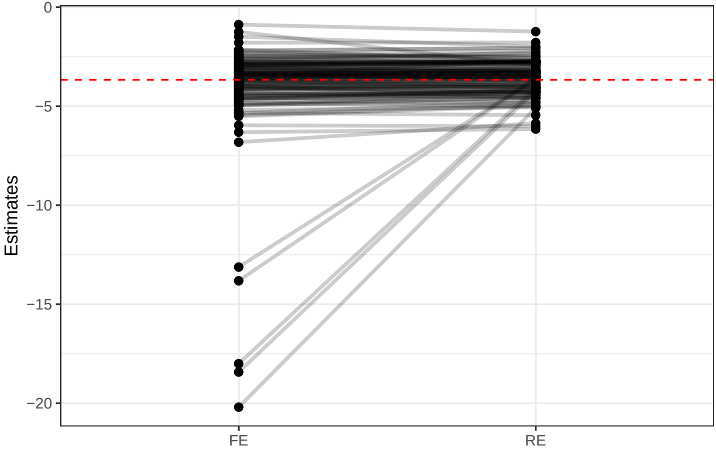
Generalized Linear Mixed-Effects Models

- Goodness-of-fit statistics are:

Model	df	LogLik	AIC	BIC
Pooled CS	2	-7449.92	14903.84	14913.13
Fixed Effects	131	-1911.19	4084.37	4692.54
Random Intercept	3	-2197.32	4400.63	4414.56
Random Coefficient	5	-2160.44	4330.87	4354.08

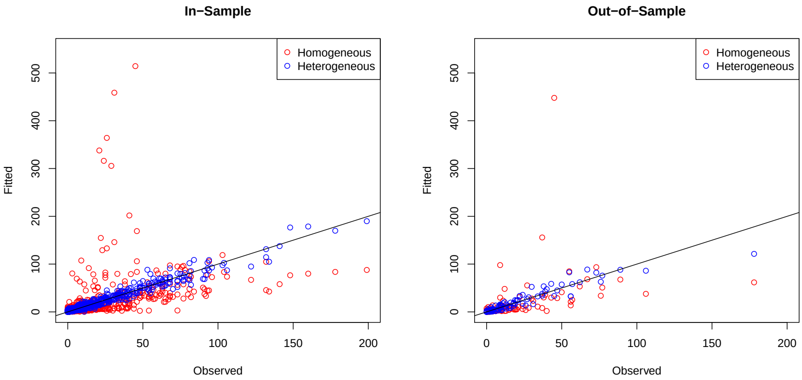
Generalized Linear Mixed-Effects Models

- Shrinkage effects:



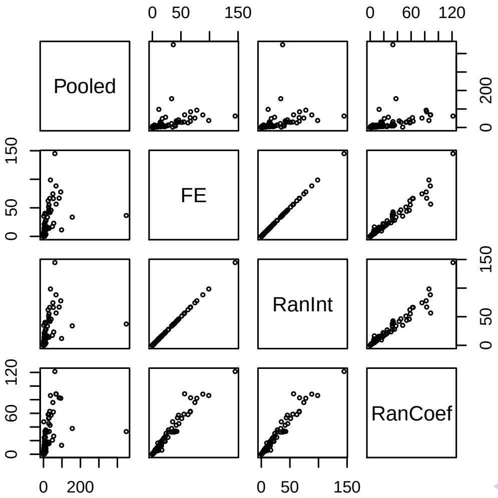
Generalized Linear Mixed-Effects Models

- Comparison of homogeneous and heterogeneous models:



Generalized Linear Mixed-Effects Models

- Comparison of out-of-sample prediction:



Generalized Linear Mixed-Effects Models

- Hold-out sample validation statistics are:

	Pooled	FE	RanInt	RanCoeff
MAE	13.281	3.885	3.871	4.443
RMSE	1.585	0.842	0.742	0.761
MAPE	41.541	6.589	6.586	8.423
Pearson	0.403	0.967	0.967	0.943
Spearman	0.821	0.936	0.938	0.945

Thank you for your attention!

References I

Klugman, S. A. (1992). *Bayesian Statistics in Actuarial Science: with Emphasis on Credibility*. Springer.