Longitudinal and Panel Data

Predictive Modeling Applications in Actuarial Science

Edward W. (Jed) Frees

University of Wisconsin – Madison

April, 2014

Outline

What are Longitudinal and Panel Data?

- With *regression data*, we collect a cross-section of subjects.
  - The interest is in comparing characteristics of the subject, that is, investigating relationships among the variables.
- In contrast, with *time series data*, we identify one or more subjects and observe them over time.
  - This allows us to study relationships over time, the so-called dynamic aspect of a problem.
- *Longitudinal/panel data* represent a marriage of regression and time series data.
  - As with regression, we collect a cross-section of subjects.
  - With panel data, we observe each subject over time.
  - Use the notation y_{it} to represent a dependent variable y observed for subject/policyholder i at time t
- The descriptor *panel data* comes from surveys of individuals; a *panel* is a group of individuals surveyed repeatedly over time.

Longitudinal/Panel Actuarial Applications

- Commercial lines insurance such as commercial auto.
  - i represents the commercial customer (policyholder) and y represents claims per premium, i.e., the loss ratio.
  - Without intervening loss reduction measures, a high loss ratio in period 1 might signal a high loss ratio in period 2.
- Insurance sales. Here, i represents a sales agent and y represents annual sales.
  - Although agent information (e.g., years of experience) and sales territory information can be useful, often prior sales history are the most important variables for predicting sales.
- Customer retention. Use $y_{i1} = 1$ to indicate that customer i bought a policy in period 1 and wish to predict y_{i2} = 1 or 0, whether or not a customer buys a policy in period 2.

Why Longitudinal and Panel Data?

- Dynamic versus Cross-Sectional Effects
  - Analysts use standard cross-sectional regression analysis to make inferences about how changes in explanatory variables will affect the dependent variable.
  - Because there is no time element in cross-sectional data, we will refer to these anticipated changes as *static*.
  - In contrast, the actuary is typically interested in changes over time, known as *temporal* or *dynamic* changes.
- Efficiency and Sharing of Information
  - Several years of information are better than one even if observations are independent
  - Common for data to exhibit features of *clustering*, where observations from the same unit of analysis tend to be similar or "close" to one another in some sense.
  - By recognizing and incorporating clustering, we can (i) develop more efficient estimators and (ii) better predictors.

Loss and Rating Variable

- To illustrate dynamic versus cross-sectional effects, consider a sample of $n = 50$ policyholders.
- This graph shows, for a single year (1), that the rating variable is an effective, although not perfect, explanatory variable for the loss.
  - As the rating variable increases, the expected loss increases.
- For year 2, suppose that a similar relationship between the loss and rating variables holds.
  - A plot of the rating variable and the loss for the combined years 1 and 2 (not pictured here) would provide the same overall conclusion as before.
- The right-hand panel shows the plot of the rating variable and the loss for the combined years 1 and 2 but with a line connecting year 1 and year 2 results for each policyholder.
  - The line emphasizes the *dynamic* effect, moving from year 1 rating variable to the year 2.

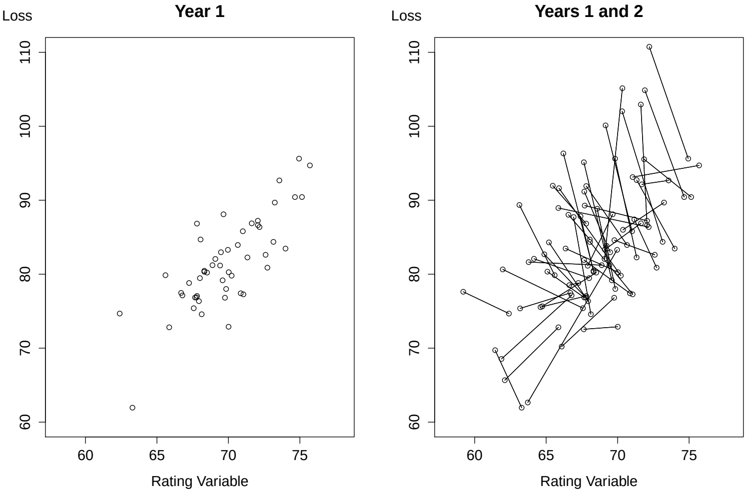

Figure : Loss and Rating Variable. The left-hand panel shows the *positive* period 1 relationship between the loss and a rating variable.
• The right-hand panel shows the loss and rating variables for both periods 1 and 2, with lines connecting the periods for each policyholder.
• Most of the lines have a *negative* slope, indicating that increases in the rating variable result in a decrease in the loss variable.

Loss and Rating Variable

- How can these data happen?
- Here are two scenarios, both driven by omitted variables that are not observable by the analyst.
- In one scenario, the rating variable naturally increases from year 1 to year 2 to due to inflation.
  - An unobserved loss reduction measure has been introduced that serves to reduce expected losses for *all* policyholders.
  - This would mean that each policyholder could expect have an increase in the horizontal x axis and a decrease in the vertical y axis, resulting in a negative slope.

Loss and Rating Variable

- How can these data happen?
- For another scenario, suppose that x represents a loss prevention measure such as a burglar alarm and y represents theft losses in a home or commercial building.
  - We might interpret the left-hand panel as both x and y being positively related to an unobserved variable such as the "safety" of the home/building's neighborhood.
  - That is, in very safe neighborhoods, theft losses y tend to be low and expenditures on burglar alarms x tend to be low (why pay a lot for a burglar alarm in such a safe neighborhood?) and conversely for unsafe neighborhoods, resulting in the overall positive slope.
  - For *each* home or building, the introduction of a more extensive burglar alarms means that expenditures x increase while expected losses y tend to decrease.
- Static analysis without paying attention to temporal effects can give a grossly biased inference about the effects of the rating variable on the losses.

Bivariate Normal Example

- Efficiency and sharing (credibility) ideas can be motivated via the normal distribution
- Suppose that we have a sample of size n of logarithmic losses from two years $\mathbf{y}_i = (y_{i1}, y_{i2})'$ that we assume are bivariate normal.
  - y_{it} is normally distributed with mean $\mu_{it} = \beta_0 + \beta_1 x_{it}$ and variance σ_t^2, for years $t = 1$ and $t = 2$, and ρ is the correlation between y_{i1} and y_{i2}.
  - x_{i1} and x_{i2} are known rating variables.
- If we want to predict year 2 losses given the information in year 1, standard probability theory tells us that the conditional distribution is normal.

$$y_{i2}|y_{i1} \sim N\left(\mu_{i2} + \rho \frac{\sigma_2}{\sigma_1}(y_{i1} - \mu_{i1}), \sigma_2^2(1-\rho^2)\right).$$

Bivariate Normal Example

- Without information about y_{i1}, the optimal predictor of y_{i2} is its mean, $\mu_{i2} = \beta_0 + \beta_1 x_{i2}$.
- With information about prior year losses, y_{i1}, we can do better.
  - The optimal predictor of y_{i2} given y_{i1} is its conditional mean, $\mu_{i2} + \rho \frac{\sigma_2}{\sigma_1} (y_{i1} - \mu_{i1})$.
  - For the conditional predictor, the stronger the relationship between the two years, the larger is the value of ρ, and the smaller is the variance of the conditional distribution.
- The conditional predictor outperforms the original (marginal) predictor because we are "sharing information" over the two years through the correlation parameter ρ.

Some Notation

- Longitudinal/panel data - regression data with "double subscripts."
- Let y_{it} be the dependent variable for the ith subject during the tth time period.
- A longitudinal data set consists of observations of the ith subject over $t = 1, \ldots, T_i$ time periods, for each of $i = 1, \ldots, n$ subjects.

$$
\begin{array}{rc}
\text{first subject} & \{y_{11}, \ldots, y_{1T_1}\} \\
\text{second subject} & \{y_{21}, \ldots, y_{2T_2}\} \\
\vdots & \vdots \\
n\text{th subject} & \{y_{n1}, \ldots, y_{nT_n}\}
\end{array}
$$

Some History

- The term “panel study” was coined in a marketing context when Lazarsfeld and Fiske (1938) considered the effect of radio advertising on product sales.
  - Traditionally, hearing radio advertisements had been thought to increase the likelihood of purchasing a product.
  - Lazarsfeld and Fiske considered whether those that bought the product would be more likely to hear the advertisement, thus positing a reverse in the direction of causality.
  - They proposed repeatedly interviewing a set of people (the “panel”) to clarify the issue.
- Baltes and Nesselroade (1979) trace the history of longitudinal data and methods with an emphasis on childhood development and psychology.
  - They describe longitudinal research as consisting of “a variety of methods connected by the idea that the entity under investigation is observed repeatedly as it exists and evolves over time.”

Data Set-Up

- There are k explanatory variables $x_{it,1}, x_{it,2}, \ldots, x_{it,k}$
- may vary by subject i and time t.
- Express the k explanatory variables as a $k \times 1$ column vector

$$\mathbf{x}_{it} = \begin{pmatrix} x_{it,1} \\ x_{it,2} \\ \vdots \\ x_{it,k} \end{pmatrix}.$$

- With this notation, the data for the ith subject consists of:

$$\begin{pmatrix} x_{i1,1}, x_{i1,2}, \ldots, x_{i1,k}, y_{i1} \\ \vdots \\ x_{iT_i,1}, x_{iT_i,2}, \ldots, x_{iT_i,k}, y_{iT_i} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{i1}', y_{i1} \\ \vdots \\ \mathbf{x}_{iT_i}', y_{iT_i} \end{pmatrix}.$$

Linear Model 1. Cross-Sectional Model

- Assume that all observations are independent, have a common variance $\mathrm{Var}\ y_{it} = \sigma^2$ and regression function

$$\begin{aligned} \mathrm{E}\ y_{it} &= \alpha + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \cdots + \beta_k x_{it,k} \\ &= \alpha + \mathbf{x}_{it}' \beta . \end{aligned}$$

- Model is regularly used as the "strawman" in any panel data analysis, one that can be easily defeated by using more sophisticated techniques.
  - Useful as a benchmark because many consumers are familiar and comfortable with cross-sectional regression analysis.
- Useful in the Capital Asset Pricing Model, known by the acronym CAPM, of stock returns.
  - In this application, i represents a firm whose stock price return, y_{it}, is followed over time t. (Sometimes it is the return in excess of the risk-free rate.)
  - The only explanatory variable is the return based on a market index.
  - Essentially, the argument from financial economics is that any patterns in the errors would be discovered, taken advantage of, and disappear, in a liquid market.

Linear Model 2. Fixed Effects Model

- In the *linear fixed effects model*, the regression function is

$$\begin{aligned} \mathrm{E}\, y_{it} &= \alpha_i + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \cdots + \beta_k x_{it,k} \\ &= \alpha_i + \mathbf{x}_{it}' \beta, \qquad t = 1, \ldots, T_i, \quad i = 1, \ldots, n. \end{aligned}$$

- The parameters $\{\beta_j\}$ are common to each subject and are called *global*, or *population*, parameters.
- The parameters $\{\alpha_i\}$ vary by subject and are known as *individual*, or *subject-specific*, parameters.
  - They often are not of primary interest - *nuisance* parameters
  - They control for differences, or “heterogeneity,” among subjects
  - Later, will be viewed as random variables, not unknown parameters

Fixed Effects Model Estimation

- Use ordinary least squares
- The heterogeneity parameters $\{\alpha_i\}$ simply represent a *factor*, that is, a categorical variable that describes the unit of observation
  - Replace categorical variables with an appropriate set of binary variables.
  - Panel data estimators are sometimes known as "least squares dummy variable model"

Analysis of Covariance Models

- There is no special relationship between subjects and time periods – we can readily interchange the roles of "i" and "t", to get

$$\mathrm{E}\, y_{it} = \lambda_t + \mathbf{x}_{it}' \beta .$$

- This model is also known as the *one-way fixed effects model*.
- Thinking of adding both factors, can readily introduce the *two-way fixed effects model*

$$\mathrm{E}\, y_{it} = \alpha_i + \lambda_t + \mathbf{x}_{it}' \beta ,$$

Linear Model 3. Random Effects Models

- In a random effects model, the variable coefficients are random variables, not fixed unknown parameters
- Now think the subjects as draws from a larger population
  - Terms such as $\{\alpha_i\}$ are draws from a distribution
  - This gives us the ability to make inferences about subjects in a population that are not included in the sample.

Random Effects Models

- The *linear random effects* model equation is

$$y_{it} = \alpha_i + \mathbf{x}_{it}' \beta + \varepsilon_{it}, \qquad t = 1, \ldots, T_i, \;\; i = 1, \ldots, n.$$

- This notation is similar to the basic fixed effects model.
- The term α_i is known as a *random effect*.
  - *Mixed effects* models are ones that include random as well as fixed effects.
  - This random effects model is a special case of the *mixed linear model*.
- We assume that $\{\alpha_i\}$ are identically and independently distributed with mean zero and variance σ_α^2.
- Further, we assume that $\{\alpha_i\}$ are independent of the disturbance random variables, ε_{it}.
- Technical Detail: Because E α_i = 0, it is customary to include a constant within the vector $\mathbf{x}_{it}$.

Random Effects Model Estimation

- Observations are no longer independent, so it is common to use *generalized least squares* estimation (GLS).
- To see the dependence, basic calculations show

$$
\begin{aligned}
\mathrm{Cov}(y_{i1}, y_{i2}) &= \mathrm{Cov}(\alpha_i + \mathbf{x}_{i1}'\beta + \varepsilon_{i1}, \alpha_i + \mathbf{x}_{i2}'\beta + \varepsilon_{i2}) \\
&= \mathrm{Cov}(\alpha_i + \varepsilon_{i1}, \alpha_i + \varepsilon_{i2}) \\
&= \mathrm{Cov}(\alpha_i, \alpha_i) + \mathrm{Cov}(\alpha_i, \varepsilon_{i2}) + \mathrm{Cov}(\varepsilon_{i1}, \alpha_i) + \mathrm{Cov}(\varepsilon_{i1}, \varepsilon_{i2}) \\
&= \mathrm{Cov}(\alpha_i, \alpha_i) = \sigma_\alpha^2.
\end{aligned}
$$

- Similarly, the variance of an observation is $\sigma_\alpha^2 + \sigma_\varepsilon^2$.
- Thus, the correlation between observations within a subject is $\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$.
- This quantity is known as the *intra-class correlation*, a commonly reported measure of dependence in random effects studies.

Linear Model 4. Model with Lagged Dependent Variables

- A direct way of relating y_{i1} to y_{i2} is through the model equation

$$y_{it} = \alpha + \gamma y_{i,t-1} + \mathbf{x}_{it}' \beta + \varepsilon_{it}.$$

- In this model equation, the dependent variable lagged by one period, $y_{i,t-1}$, is used as an explanatory variable to predict y_{it}. The parameter γ controls the strength of this relationship.
- A strength of this model is that it is easy to interpret and to explain.
- It is similar in appearance to the popular autoregressive model of order one, $AR1$.
  - With the $AR1$ models, one loses the time $t = 1$ observations because there is no lagged version for the first period.
  - However, for panel data, this means losing n observations.

Linear Model 5. Models with Serial Correlation

- Time trends are important in longitudinal data, we need a variety of tools to accommodate different dynamic patterns
- Incorporating year or other time dependent fixed effects is one way
- Another possibility is to use a lagged dependent variable as a predictor.
  - However, this is known to have some unexpected negative consequences for the basic fixed effects model (see for example the discussion in Hsiao, 2003, Section 4.2 or Frees, 2004, Section 6.3).
- Another approach, examine the serial correlation structure of the disturbance term $\varepsilon_{it} = y_{it} - \mathrm{E}\ y_{it}$.
- For example, a common specification is to use an autocorrelation of order one, $AR(1)$, structure, such as

$$\varepsilon_{it} = \rho_{\varepsilon} \varepsilon_{i,t-1} + \eta_{it},$$

where $\{\eta_{it}\}$ is a set of disturbance random variables and ρ_{ε} is the autocorrelation parameter.

Example: Group Term Life

- Claims and exposure information from 88 Florida credit unions for years 1993-1996.
- "Life savings" claims from a contract between the credit union and their members that provides a death benefit based on the member's savings deposited in the credit union.
  - Actuaries typically price life insurance coverage with knowledge of an insureds' age and gender, as well as other explanatory variables such as occupation.
  - However, for these data from small groups, often only a minimal amount of information is available to understand claims behavior.
- Of the $88 \times 4 = 352$ potential observations, 27 were not available because these credit unions had zero coverage in that year (and thus excluded). Thus, these data were unbalanced.

Summary Statistics – Group Term Life Data

- The dependent variable is the annual total claims from the life savings contract.
- The explanatory variables is the annual coverage.
- It turns out that both coverages and claims are highly skewed and so we will analyze their (natural) logarithmic versions.
- We use the transformation $\texttt{LnClaims} = \ln(\texttt{1+Claims})$, so that credit unions with zero claims remain at zero when on a logarithmic scale (and similarly for coverages).

	Mean	Median	Standard Deviation	Minimum	Maximum
Coverage (000's)	30,277	11,545	54,901	25	427,727
Claims (000's)	14.724	5.744	32.517	0	290.206
Logarithmic Coverage	16.272	16.262	1.426	10.145	19.874
Logarithmic Claims	8.029	8.656	2.710	0	12.578

Visualizing Group Term Life Claims over Time

- To visualize the claim development over time, we provide a *trellis plot* of logarithmic claims.
- This plot shows logarithmic claims versus year, with a panel for each credit union, arranged roughly in order of increasing size of claims.
- A trellis plot provides extensive information about a data set and so is unlikely to be of interest to management although can be critical to the analyst developing a model.
- The plot shows that claims are increasing over time and that this pattern of increase largely holds for *each* credit union.
  - Credit union number 26, in the upper right hand corner, has substantially larger claims than even the next largest credit union.
  - As the typical size of the claims decreases, the (downward) variability increases.
  - Much can be learned by the analyst from close inspection of trellis plots.

Visualizing Group Term Life Claims over Time

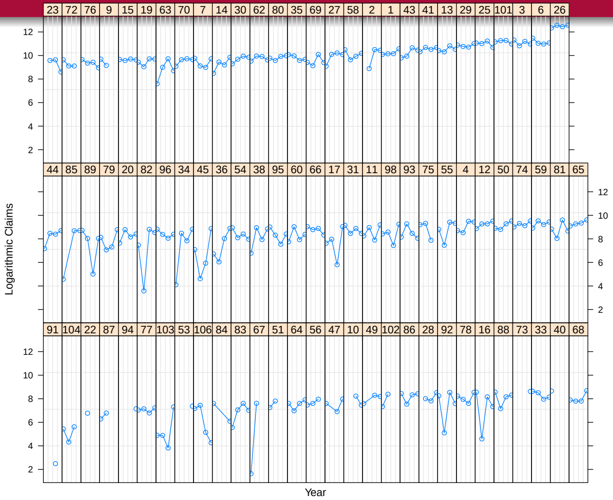

Figure : Trellis Plot of Logarithmic Group Term Life Claims over 1993-1996.

Summarizing Five Model Fits

- Interestingly, the coefficient estimates across the five models are relatively consistent, indicating that the parameter estimates are robust to model specification.
- For the lagged dependent variable model, the intercept is $\frac{-8.8657}{1-0.303} = -12.420$, consistent with the other models.
- Additional examination of model diagnostics (not displayed here) shows that all four longitudinal data models are superior to our "strawman," the ordinary cross-sectional regression model.

	Models				
	Cross Sectional	Fixed Effects	Random Effects	Lagged Dependent	Correlated Errors
Intercept	-12.337	-12.286	-12.882	-8.657	-12.567
t-statistic	-9.49	-1.65	-7.48	-5.80	-7.73
Logarithmic Coverage	1.252	1.264	1.282	0.884	1.265
t-statistic	15.73	3.04	12.14	8.36	12.69
Lagged Claims				0.303	
t-statistic				5.35	

Non-Linear Models

- Although the linear model framework provides a convenient pedagogic framework to base our discussions, many actuarial applications fall into the non-linear model context.
- This section provides a road map of how to think about modeling choices when your data can not be reasonably be represented using a linear model.

Binary Outcomes

- Many actuarial applications involve analyses of data sets where the outcome of interest is binary, often using $y = 1$ to signal the presence of an attribute and $y = 0$ to signal its absence.
- For a policy, this can be a check function to see whether or not there was a claim during the period.
- For a customer, this can be whether or not a customer from one period is retained from one period to the next.

Binary Outcomes with Random Effects

- It is common to use a random effects model to reflect clustering of binary longitudinal data.
- To see how to incorporate random effects, use a logistic function $\pi(z) = \ln \frac{1}{1+e^{-z}}$.
- Now, conditional on α_i, define the probability

$$\pi_{it} = \Pr(y_{it} = 1 | \alpha_i) \quad = \quad \pi(\alpha_i + \mathbf{x}_{it}' \beta).$$

- As with linear random effects models, the quantity α_i can capture effects for the ith subject that are not observable in the other explanatory variables.
- Estimation of binary outcomes random effects models is generally conducted using maximum likelihood.

Binary Outcomes and Markov Transition Models

- In actuarial applications, it is helpful to think about Markov transition modeling.
- With these models, actuaries can account for *persistency* by tracing the development of a dependent variable over time and representing the distribution of its current value as a function of its history.
- Think about the events $\{y = 1\}$ and $\{y = 0\}$ as representing two "states." "Persistency" connotes the tendency to remain in a state over time.
- It is the same idea as clustering yet applied to a state space context.

Binary Outcomes and Markov Transition Models

- For a Markov model (of order 1), one assumes that the entire history H is captured by the most recent outcome, so that

$$f(y_{it}|H_{it}) = f(y_{it}|y_{i,t-1}).$$

- For binary outcomes, we can write the conditional probability of a 1, given $y_{i,t-1} = 0$, as

$$\pi_{it,0} = \Pr(y_{it} = 1|y_{i,t-1} = 0) \quad = \quad \pi(\alpha_0 + \mathbf{x}_{it}'\beta_0)$$

and, given $y_{i,t-1} = 1$, as

$$\pi_{it,1} = \Pr(y_{it} = 1|y_{i,t-1} = 1) \quad = \quad \pi(\alpha_1 + \mathbf{x}_{it}'\beta_1).$$

- In this context, $\pi_{it,0}$ and $\pi_{it,1}$ are examples of *transition probabilities*, quantifying the probability of moving or transiting from one state to another.

Binary Outcomes and Markov Transition Models

- With this notation, the conditional distribution is given as

$$f(y_{it}|y_{i,t-1}) = \begin{cases} \pi_{it,1} & \text{if } y_{i,t-1}=1, y_{it}=1 \\ 1-\pi_{it,1} & \text{if } y_{i,t-1}=1, y_{it}=0 \\ \pi_{it,0} & \text{if } y_{i,t-1}=0, y_{it}=1 \\ 1-\pi_{it,0} & \text{if } y_{i,t-1}=0, y_{it}=0 \end{cases}.$$

  Then, it is customary to estimate model parameters by maximizing a *partial log-likelihood*, given as

$$L_P = \sum_i \sum_{t=2}^{T_i} \ln f(y_{it}|y_{i,t-1}).$$

- As with lagged dependent variable linear model, this modeling choice does lose the period 1 observations; in this sense, it is a "partial" likelihood. For many problems of interest, one loses little by focusing on the partial likelihood.

Generalized Linear Model Outcomes and Random Effects

- Many "non-normal" outcomes can be handled using a generalized linear model (GLM).
- To handle clustering in a panel data context, random effects are commonly used.
- For this formulation, we follow the usual three-stage GLM set-up. Specifically, we
  - Specify a distribution from the linear exponential family of distributions.
  - Introduce a systematic component. With random effects, this is conditional on α_i so that $\eta_{it} = \alpha_i + \mathbf{x}_{it}^{\prime} \beta$.
  - Relate the systematic component to the (conditional) mean of the distribution through a specified link function

$$\eta_{it} = g(\mu_{it}) = g\left(\mathrm{E}\left(y_{it} | \alpha_i\right)\right).$$

Generalized Linear Model Outcomes and Random Effects

- Many "non-normal" outcomes can be handled using a generalized linear model (GLM).
- This modeling framework is sufficiently flexible to handle many practical applications.
- From a user's viewpoint, it is convenient to have a single statistical software program regardless of whether one wants to model a count (e.g., Poisson) or a medium-tailed (e.g., gamma) distribution.
- The random effects GLM models have the same limitations discussed in the special case of binary outcomes.
  - It is a computationally intensive formulation that is tractable only with modern-day software and hardware.
  - A fixed effects version is often not a reliable alternative.

Additional Considerations

- Unbalanced and Missing Data
- Clustered Data (non-temporal)

Additional Resources

- Actuarial Examples
  - Three examples are in the paper
  - Frees, Edward W., Virginia R. Young and Yu Luo (2001). Case studies using panel data models. *North American Actuarial Journal* 5 (4), 24-42.
  - Data and SAS code available at ,

    `http://instruction.bus.wisc.edu/jfrees/jfreesbooks/Longitudinal%20and%20Panel%20Data/Book/PDataBook.htm`
- The ideas in this chapter are expanded upon in the book-length treatment of Frees (2004).
  - Book info at www.cambridge.org,

    `http://www.cambridge.org/gb/knowledge/isbn/item1170984/?site_locale=en_GB`
  - Book web site

    `http://instruction.bus.wisc.edu/jfrees/jfreesbooks/Longitudinal%20and%20Panel%20Data/Book/PDataBook.htm`

Book URL

- You can learn more about the book at the Cambridge University Press website

  `http://www.cambridge.org/us/academic/subjects/statistics-probability/statistics-econometrics-finance-and-insurance/predictive-modeling-applications-actuarial-science-volume-1`

- Book Resources (data, sample code) are available at `http://research.bus.wisc.edu/PredModelActuaries`