ACTEX Learning

Study Manual for ILA Life ALM and Modeling Exam

3rd Edition

John Aprill, FSA, MAAA

An SOA Exam

ACTEX Learning

Learn Today. Lead Tomorrow.

Actuarial & Financial Risk Resource Materials

Since 1972

Printed in the United States of America.

TABLE OF CONTENTS

Section A: Stochastic Models, Generalized Linear Models, Multi-State and Transition Matrix Models

Section B: Issues Common to Models (Including Model Governance and Data Issues)

<u>Section C: Asset-Liability and Projection Models</u>

Section D: Economic Scenario Generators and Equity Linked Models

Section E: Asset and Portfolio Management Topics

SECTION A

STOCHASTIC MODELS, GENERALIZED LINEAR MODELS, MULTI-STATE AND TRANSITION MATRIX MODELS

HANDBOOK OF FIXED INCOME SECURITIES, CH 49

INTRODUCTION TO MULTIFACTOR RISK MODELS IN FIXED INCOME AND THEIR APPLICATIONS

I. Motivation and structure underlying fixed income multifactor risk models

A. Risk management is part of the investment process; risk models allow managers to quantify and analyze the risk embedded in their portfolios

1. Risk models give managers insight into the source of risk in a portfolio and help them control their exposures and understand contributions of different portfolio components to total risk

2. Risk models help managers in the decision-making process as to how to make changes to the portfolio

B. One method to analyze portfolio volatility is to compute the volatility of historical returns and use this measure to forecast future volatility; however, this framework does not provide insight into

1. Relationships among securities or the major sources of risk

2. Changes to the characteristics of a fixed income portfolio over time as instruments mature were subject to credit events

C. Portfolio volatility can utilize a different strategy

1. The portfolio return is a function of individual instrument returns and the market weights of the securities in the portfolio

2. The forecasted volatility of the portfolio σ_P^f can be computed as a function of these weights w and the covariance matrix Σ_S of the instrument returns in the portfolio

$$\sigma_P^f = \text{sqrt}\,(w^T \times \Sigma_S \times w) \quad \text{where } T \text{ denotes the matrix transpose}$$

3. This covariance matrix can be decomposed into the individual instrument volatilities and correlations among returns

4. Using this approach, the portfolio manager gains insight into the riskiness of the portfolio and how the portfolio can be diversified

a. Portfolio volatility requires estimates of correlations among each pair of instruments which grows quadratically with the number of instruments in the portfolio and prevents estimation of the relationship between turns in a robust way

b. This approach also uses historical returns to calculate future volatility, but these can change from time to time

D. Multi-factor models

1. A major characteristic of multi factor models is their ability to provide a return to the portfolio using a small set of variables called factors which are designed to capture broad systematic market fluctuations and the nuances of individual portfolios

2. Most factor models are linear in that the total return can be decomposed into the sum of the contributions of the factors

   a. Systematic return is a component of total return due to movements in the common risk factors

   b. Idiosyncratic return can be described as a residual component that cannot be explained by systematic factors systematic factors

3. Correlations across securities of different issuers are driven by their exposures to the systematic risk factors and the correlation between those factors

4. The equation below demonstrates the systematic and idiosyncratic components of total return for a security *s* where the systematic return is a product of the instrument's loadings to the systematic risk factors *F* and the returns of these factors, and the idiosyncratic return is ε_s

$$R_s = L_s \times \mathrm{F} + \varepsilon_s$$

5. Under these models of portfolio volatility can be estimated as

$$\sigma_P^f = \mathrm{sqrt}\,(L_p^T \times \Sigma_F \times L_p + w^T \times \Omega_S \times w)$$

   where

   L_p is the loading to the portfolio to the risk factors
   Σ_F is the covariance matrix of factor returns
   Ω_S is the covariance matrix of the idiosyncratic security returns which are assumed to be uncorrelated

6. The idiosyncratic return of securities is diversified away as the number of securities in the portfolio increases (the diversification benefit)

7. A more robust estimation is created because the number of parameters is smaller in Σ_F is smaller than Σ_s

8. Factors can be designed to be more stable than individual stock returns leading to models with better predictability

9. An important advantage of using multi-factor models is the insight they provide into the structure and properties of the portfolio and enable decision making

10. The model success depends on the ability to interpret historical and current data to formulate estimates of future portfolio risk; therefore, the properties of the data used should produce an error in the estimate of future realizations that is relatively small

11. Even sophisticated models are subject to event risk caused by geopolitical or financial events which are followed by a period of

    a. Large negative returns of risky assets
    b. Large positive returns of assets considered safe and
    c. Significantly higher volatility and very high correlations

12. Event risk will not be captured in the projections and must be tested using a "what if" analysis under stress conditions

II. Fixed income risk models

A. Systematic risk factors

1. Types of risk for fixed income securities

    a. Systematic fixed income risk factors are

        i. Risks with influence across asset classes (yield curve)
        ii. Sector specific to a particular asset class (prepayment)

    b. All risk not captured by systematic factors is idiosyncratic risk

    c. Systematic risk factors can be defined by purely statistical methods observed in the market or estimated from asset returns; pricing models can also be used

2. Global risk model

    a. Risk factors are other observable or estimated from regressing cross-sectional asset returns on instrument sensitivities

    b. The forecasted systematic risk is a function of the mismatch between the portfolio and the benchmark, and the exposure is to the risk factors such as spreads

    c. Net portfolio exposures are aggregated from security level analytics

    d. Systematic risk is also a function of volatility of risk factors and the correlations between risk factors

    e. Because the model uses security level returns and analytics to estimate the factors, it can recover the idiosyncratic return for each security

3. Curve risk

    a. Curve risk is the major source of risk across fixed income instruments; this matches curve profiles relative to a benchmark and is usually the main driver of portfolio risk

    b. Typically, the benchmark is the government or the swap curve

        i. During calm periods the behavior of the swap curve tends to match the government curve

        ii. However, during a liquidity crisis they can diverge significantly

    c. For government-based products the government curve is used; for all other products the spreads between the swap and the government curve are used

    d. The risk associated with each of these curves is the exposure of the portfolio at different points along the curve combined with volatility and correlation of the movement of the curve points

    e. Sometimes a convexity term is required to capture second order exposure to curve changes on assets with long tenors or embedded optionality

    f. The statistical method used most often is a principal component analysis

        i. The method defines factors statistically independent of each other

        ii. Generally, three or four sectors are sufficient to explain the risk associated with changes across the yield curve

    g. Portfolio managers favored the key rate approach because

        i. It uses a larger set of economic factors and

        ii. It is more intuitive and captures the risk of specialized portfolios better

4. Credit risk

    a. Bonds issued by corporations have credit risk, and holders of these securities demand additional yield above the risk-free rate to compensate for that risk

    b. The risk is usually measured by reference to a curve such as a swap curve; the total of credit spreads determines the credit risk exposure associated with the portfolio

    c. Characteristics of credit bonds have systematic sources of credit spread risk such as the industry, credit ratings and the country of the issuer

    d. The loading of a bond to a credit risk factor would commonly use spread duration multiplied by the bonds spread

    e. Specific risk factors are not needed; it is this automatically incorporated into the bond loading in the risk factor

5. Prepayment risk

    a. Securitized products are exposed to prepayment

    b. The most common are residential mortgage-backed securities which represent pools of borrowers who may prepay their debt before maturity when prevailing lending rates are lower

    c. The risk to the security holder is that they may hold cash when reinvestment rates are low

    d. Therefore, these securities have two sources of risk: interest rate (including convexity) and prepayment risk

    e. Part of the prepayment risk can be expressed as a function of interest rates using a prepayment model

        i. The model captures interest rates using key rate durations and their convexity

        ii. Convexity is generally negative and has a detrimental effect on the market value of the instrument when interest rates move in either direction

            a) Decreasing interest rates cause prepayments to increase thereby reducing price appreciation due to falling rates

            b) Rising interest rates intensify the price depreciation the instrument suffers with the higher rates

f. The remaining prepayment risk must be modeled with additional systematic risk factors

g. Prepayment spreads on mortgage-backed securities depend on the program and term of the deal, whether the bond is priced at a discount or a premium, and how seasoned the bond is

6. Implied volatility risk

   a. Many fixed income securities have embedded options for interest rates or the discount curve used to price the security; if expected volatility increases, options generally become more expensive thereby affecting the prices of bonds with embedded options

   b. Exposure to implied yield curve volatility is also a source of risk

   c. Sensitivity of securities to changes in implied volatilities is measured by vega, calculated from the security pricing model

   d. Implied volatility factors can be calculated from

      i. The market price of liquid fixed income options or
      ii. The implied volatility by the returns of bonds with embedded options in each asset class

7. Liquidity risk

   a. Fixed income securities traded in decentralized markets are illiquid, making it difficult to establish a fair price

   b. These bonds are exposed to liquidity risk because the illiquid bondholder may have to pay a higher price to liquidate a position, generally selling at a discount

   c. The amount of the liquidity discount is uncertain and changes during the business cycle

8. Inflation risk

   a. Inflation securities are based on expectation of future inflation

   b. This uncertainty adds to the volatility of a bond over the volatility of other risk sources

   c. Expected inflation is not an observable variable in the marketplace but can be extracted from prices of inflation linked government bonds and inflation swaps

   d. The sensitivity of securities to expected inflation uses a specialized pricing model and is called inflation duration

9. Tax policy risk

    a. Many securities are tax exempt which provide an additional benefit based on the allowable tax exemption that is incorporated in the price of the security

    b. Uncertainty around tax policy adds to the risk of these securities

    c. This cannot be observed in the marketplace and must be extracted from the prices of municipal securities

    d. The return on municipal securities in excess of interest rates is driven by tax policy expectations and the creditworthiness of the issuer which can be difficult to separate

B. Idiosyncratic risk

1. Once all systematic factors and holdings are determined, the residual idiosyncratic return of a security can be computed as the component of its total return that cannot be explained by systematic factors

2. Idiosyncratic return can be a significant component of total return but tends to decrease rapidly as the number of instruments increases

3. The major inputs to idiosyncratic risk are the instrument characteristics and historical idiosyncratic returns of the instruments, such as instrument spread, spread duration, industry membership, and idiosyncratic volatility

4. Idiosyncratic returns of different issuers are considered to be uncorrelated; however, different securities from the same issuer can have a level of co-movement

III. Applications of risk modeling

A. All risk models translate portfolio characteristics across different dimensions into a common, comparable set of numbers

B. The investment process starts with an investment universe and objectives and has several stages: portfolio construction, risk production, and performance evaluation

C. Portfolio construction and risk budgeting

1. During the portfolio construction exercise, the manager must carefully choose exposures to risk factors to achieve the highest possible portfolio return subject to risk and other constraints

2. Risk models help managers achieve this goal in an objective and quantifiable way

3. Indexers Are required to follow a given benchmark index with the minimum possible deviation to avoid tracking error and to minimize portfolio costs

4. Enhanced indexers are allowed some leeway to deviate from a benchmark to achieve superior returns

    a. The leeway can be expressed by a set of rules constraining exposure to risk factors (often prescribed by a risk budget)

    b. A risk budget is an upper limit statistical measure of deviation between the portfolio and the benchmark

    c. Tracking error volatility (TEV) of the return of the portfolio net of the benchmark is the most common measure of deviation

5. Absolute return portfolio managers do not track to a benchmark but aim for the highest possible portfolio return which may be subject to leverage, portfolio composition, exposure, and risk constraints

D. Analyzing portfolio risk

1. The main application of risk models is the risk measurement of portfolios relative to a benchmark

2. Risk analysis depends on multi factor models either in the aggregate or in depth with each individual instrument

3. A multi factor approach can analyze portfolio risk in a number of dimensions, including risk factor exposure, contributions to total risk, and the analysis of the issuer level

4. Risk can be composed into systematic and idiosyncratic risks which are independent because they constitute isolated and uncorrelated risk sources of the portfolio

5. An alternative method to describe how different risk sources impact overall portfolio risk is risk attribution

$$Total\ TEV = \sqrt{Systemic\ TEV^2 + Idiosyncratic\ TEV^2}$$

   a. Risk attribution allows a portfolio manager to decompose portfolio volatility in an additive manner

   b. It can be used to decompose portfolio risk into contributions from different buckets representing securities or risk factors

   c. Many additional risk analytics can be computed on the basis of a linear factor model such as betas to the benchmark, liquidation effect, and marginal contributions to risk

E. Portfolio rebalancing

1. Most managers rebalance their portfolios at regular intervals to reflect changing views and market circumstances

2. Over time, statistics drift from targeted levels due to aging of the holdings, changes in the market environment, or issuer specific events

3. A risk model is very useful in rebalancing a portfolio

4. A risk model can tell the manager how much risk reduction a transaction can achieve in order to evaluate the risk reduction relative to transaction cost

5. Rebalancing and extra constrains imposed on the optimization problem may not materially change the risk of the portfolio

F. Scenario analysis

1. A portfolio manager expresses his opinion on certain financial variables and risk factors to construct a scenario

2. Not all risk factors will be known or have known behavior

3. These partial views can be completed using statistical assumptions and a covariance matrix

4. Mechanics for a scenario return of the portfolio

   a. The manager translates his views into risk factors

   b. The covariance matrix is used to complete all risk factors

   c. Net loadings of all risk factors are used to get the net return under the scenario

5. The assumptions

    a. The manager can represent his views as risk factor returns

    b. To complete a scenario a stationary and normal multivariate distribution among all factors is assumed

6. Factor based scenario analysis can significantly increase the intuition of the portfolio manager when using risk models

7. If a benchmark has a lower spread than the risk model, it will have higher profits in a scenario

IV. Key points

A. Risk models describe different imbalances of a portfolio using a common language; imbalances are combined into a consistent and coherent analysis reported by the risk model

B. Risk models provide important insights regarding different trade-offs existing in the portfolio and guidance as to how to balance them

C. Fundamental systematic risk in a fixed income security portfolio is interest rate and term structure risk that are captured by factors representing risk free rates and swap spreads of different maturities

D. Excess systematic risk over interest rates is captured by factors specific to each asset class including credit risk, prepayment risk, volatility, liquidity, inflation, and tax policy

E. Idiosyncratic risk is diversified away in large portfolios but can become a significant component in a small portfolio; correlation of idiosyncratic risk or securities of the same issuer must be modeled carefully

F. A good risk model provides detailed information about the exposures of a complex portfolio and it is a valuable tool for portfolio construction and management

    1. Managers can construct portfolios tracking a benchmark, express views of a given risk budget, and rebalance a portfolio without excessive transaction costs

    2. By identifying the exposures in a portfolio with highest risk sensitivity, the portfolio manager can effectively reduce or increase risk

**STOCHASTIC MODELING, CHAPTER I,
GENERAL METHODOLOGY**

I. Actuarial science is the quantification, analysis and management of future contingent risk and its financial consequences; it involves making assertions about quantities where actual values are uncertain

II. Stochastic models vs. non-stochastic models

A. A stochastic model

1. Is a mathematical simplification of a process involving random variables
2. The primary purpose is to provide a projection based on a range of a single set of assumptions selected by the user
3. It may be average expected outcomes or stress scenarios

B. When should stochastic models be used?

1. When required by regulation or standards of professionalism
2. When analyzing extreme outcomes or tail risks that are not well understood
3. When using certain risk measures such as Value at Risk (VaR) or conditional tail expectation (CTE)
4. When certain percentiles are required
5. When one wants to understand where stress tests fall in the spectrum of possible outcomes

C. When should use of stochastic models be questioned?

1. When it is difficult or impossible to determine the appropriate probability distribution
2. When it is difficult or impossible to calibrate the model
3. When it is difficult or impossible to validate the model

D. Alternatives to stochastic models

1. Stress testing or scenario testing
2. Static factors or load factors
3. Ranges

E. Disadvantages of stochastic models

1. The black box phenomenon
2. Improper calibration or validation
3. Uses of inappropriate distributions or parameters

F. Guidance on stochastic model implementation

1. Describe the goals and intended uses of the model
2. Decide if stochastic modeling is necessary or if an alternative approach will yield equally useful results
3. Decide on the risk metrics
4. Establish which risk factors need to be modeled stochastically
5. Determine the approach for modeling these risk factors in terms of which distributions or models should be used and how to parameterize or fit the distributions
6. Determine the number of scenarios necessary to reach the point at which additional iterations provide no additional information about the shape of the distribution
7. Calibrate the model
8. Run the model
9. Validate the model and review output
10. Conduct a peer review
11. Communicate results

III. Risk neutral vs. real world

A. Risk neutral scenarios

1. Uses

a. What is the market consistent value or fair value of an insurance liability?
b. What is the expected hedging cost of an insurance guarantee with an embedded derivative?
c. What is the price for fair value of an exotic derivative?
d. How much would a market participant demand to assume liability cash flows?

B. Real world scenarios

1. Background

a. Seek to maintain consistency with stylized facts including

a. Risk premiums above the risk free rate that equities earn over the long term
b. Term premiums that investors in longer term bonds require over yields of shorter term bonds
c. Credit spreads in excess of long term default costs that compensate the holders of credit risk
d. Implied volatilities reflected in option prices that are in excess of realized return volatility exhibited by equities; implied volatilities are commonly measured using market observed prices and a pricing model such as Black-Scholes

2. Uses

   a. What does the distribution of the present value of earnings look like?
   b. How much capital does the company need to support worst case scenarios?
   c. What level of pricing is required to earn a target return of at least x% of the time without assuming excessive downside risk?
   d. What kind of earnings volatility could a block of business generate?
   e. How does the distribution of final capital change with changes in investment strategy?
   f. How much residual risk remains from alternative hedging or risk management strategies?

C. Deflators

1. Practical considerations

   a. Deflators tend to be numerically unstable as the growth rate and risk free rate are simulated at the same time

   b. Deflators tend to require more simulations for convergence than risk neutral valuations

   c. Deflators require additional sophistication of the economic scenario generator because

      a. The need to find a different set of deflators for each real world parameterization
      b. The need to find a different set of deflators for each asset modeled
      c. The need to explore further complications for items such as stochastic volatility

   d. Deflators tend to be less transparent than risk free rates

   e. Because of incomplete markets,

      a. It is not possible to replicate the value of some assets
      b. There is often more than one possible risk neutral measure
      c. Risk neutral and deflator models can give different, arbitrage free results
      d. For path dependent options, further pricing differences can occur compared to risk neutral valuations

LAM-137-19
MULTI-STATE TRANSITION MODELS WITH ACTUARIAL APPLICATION

I. Multi-state transition models for actuarial applications

A. Introduction

1. Subject leaves a state and cannot return to the state

a. Example: basic survival model

a. One state—life aged x fails at time $T(x)$
b. Two states—status is either alive or failed

b. Example: multi-decrement survival models
The time of failure and the cause of failure is important

c. Example: multi-life models; failure occurs on the order of death

2. Subject leaves a state but can return to the state

a. Example: disability

a. Statuses: active, temporarily disabled, permanently disabled and inactive
b. Models describe the probability of moving between active and temporarily disabled

b. Example: driver ratings

a. Statuses: preferred, standard, substandard
b. Models describe the probability of moving among these ratings

c. Example: continuing care retirement communities (CCRCs)

a. Statuses: independent living, temporarily in health center, permanently in health center and gone
b. Models describe the probability of moving among these states

B. Non-homogeneous Markov chains

1. Simplifying assumptions

a. Discrete time; states described at time 0, 1, …

b. A finite number of states

c. History independence—the probability distribution at time $n+1$ depends on time n and on the state of n, but does not depend on the states at times prior to n

2. Definition: M is a non-homogenous Markov chain when M is an infinite sequence of random variables M_0, M_1,… with the following properties

a. M_n denotes the state number of a subject at time n

b. Each is a discrete type random variable over r values (either 1, 2, …, r or 0 1, …, m with $r = m + 1$)

The transition probabilities shown below are history independent

$$Q_n^{(i,j)} = \Pr[M_{n+1} = j \mid M_n = i \text{ and previous values of } M_k] = \Pr[M_{n+1} = j \mid M_n = i]$$

3. If the transition probabilities $Q_n^{(i,j)}$ depend on n, they are denoted $Q^{(i,j)}$ and the chain is a homogenous Markov Chain

C. More probabilities

1. Notation: $P_n^{(i)} = Q_n^{(i,i)}$ is the success probability of remaining in State #i

2. Definition (transition probability matrix):

The transition probability matrix $\boldsymbol{Q}_n$ is the r by r matrix whose entry in row i and column j--the (i,j) entry—is the transition probability $Q_n^{(i,j)}$

3. Notation:

$Q_n^{(i,j)} = \Pr[M_{n+k} = j \mid M_n = i]$ with ${}_k\boldsymbol{Q}_n$ used by the r by r matrix whose (i,j) entry is ${}_kQ_n^{(i,j)}$

4. Theorem (longer term probabilities):

In non-homogenous Markov chains, the longer term probability can be computed as the (i,j) entry of the matrix $\boldsymbol{Q}_n \times \boldsymbol{Q}_{n+1} \times \ldots. \times \boldsymbol{Q}_{n+k-1}$

$${}_k\boldsymbol{Q}_n = \boldsymbol{Q}_n \times \boldsymbol{Q}_{n+1} \times \ldots. \times \boldsymbol{Q}_{n+k-1}$$

For a homogenous Markov chain, the matrix is $\boldsymbol{Q}^k$

5. Theorem: The probability that a subject in state #i at time n remains in that state through time $n + k$ is

$$_{k}P_{n}^{(i)} = P_{n}^{(i)}\, P_{n+1}^{(i)} \ldots\, P_{n+k-1}^{(i)}$$

6. Theorem (future transition probabilities): Given that a subject is in state #s at time n, the probability of making the transition from state #i at time $n + k$ to state #j at time $n + k + 1$ is

$$_{k}Q_{n}^{(s,i)}\, Q_{n+k}^{(i,j)}$$

II. Cash flows and their actuarial present values

A. Introduction

1. Actuaries need cash flow models associated with future states

2. Example: insurance and annuities

a. Payments made upon failure of a status (life insurance) or payments made while a status is intact (annuities)
b. In the non-homogenous Markov chain model, insurance payments model life insurance payments that correspond to payments made from one state to another and annuities represent payments made while the subject is in a particular state

3. Example: disability

a. Concern is payments made to an employee while temporarily or permanently disabled and administrative costs associated with a change of status
b. These correspond to cash flows while the subject is in a particular state or on transition from one state to another

4. Example: driver ratings

a. Concerns are the expected claims payable and premiums collected while a driver is in a particular classification
b. In non-homogenous Markov chain models, these correspond to cash flows while the subject is in a particular state or on transition from one state to another

5. Example: CCRCs

a. Concerns are expenses to be paid and payments collected while a resident is in a particular classification
b. In non-homogenous Markov chain models, these correspond to cash flows while the subject is in a particular state or on transition from one state to another

B. Cash flows upon transition

1. Notation (cashflows at transition):

$_{t+1}C^{(i,j)}$ denotes the cash flow at time $\ell + 1$
if the subject is in state #i and state #j at time $\ell + 1$

2. Notation (discounting): $_{k}v_{n}$ denotes the value at time n of one unit certain to be paid k periods in the future at time $n + k$

C. Actuarial present values

1. Triple product summation (the 3πΣ approach) is the product of three terms

   a. The probability that the cash flow occurs
   b. The amount of the cash flow
   c. The discounting from the time of the cash flow back to the present time (n)

2. For a subject in state #s at time n

   a. Cash flows can occur at times $n + k + 1$ for $k \geq 0$

   b. The amount of the cash flow at that time is denoted by $_{n+k+1}C^{(i,j)}$

   c. The discounting from the time $n + k + 1$ to present time n is $_{k+1}\text{v}_{\text{n}}$

   d. The probability of making the transition from State #i at time $n + k$ to State #j at time $n + k + 1$ is $_{k}Q_{n}{}^{(s,i)}\ Q_{n+k}{}^{(i,j)}$

3. Theorem--actuarial present value of cash flows at transitions:

   a. Let $_{t+1}C^{(i,j)}$ denotes the cash flow at time $\ell + 1$
   if the subject is in state #i and state #j at time $\ell + 1$

   b. Assume the subject is now in state #s at time n

   c. Then the actuarial present value at time n of these cash flows given by the triple product summation (3πΣ)

$$APV_{s@n}\ (C^{(i,j)}) = \sum_{\mathbf{k=0}}^{\infty} [\ {}_{k}Q_{n}{}^{(s,i)}\ Q_{n+k}{}^{(i,j)}\ ]\ [\ {}_{n+k+1}C^{(i,j)}\ ]\ [\ {}_{k+1}v_{n}\ ]$$

D. Cash flows while in states

1. Cash flows can now occur while the subject is in a state rather than on transition between states

2. In computing the triple product summation (3πΣ), the difference is in the probability that the cash flow occurs

3. Notation cash flows while in states):

   ${}_{\ell}C^{(i)}$ denotes the cash flow at time ℓ
   if the subject is in state #i and state #i at time ℓ

4. Triple product summation (the 3πΣ approach) is the product of three terms

   a. The probability that the cash flow occurs

   b. The amount of the cash flow

   c. The discounting from the time of the cash flow back to the present time (n)

5. For a subject in state #s at time n

   a. The times the cash flow can occur are times $n + k$ for $k \geq 0$

   b. The probability of being in state #i at that time is ${}_{k}Q_{n}^{(s,i)}$

   c. The amount of the cash flow at that time is denoted by ${}_{n+k}C^{(i)}$

   d. The discounting from the time $n + k$ to present time n is ${}_{k}v_{n}$

6. Theorem (actuarial present value of cash flows while in states):

   a. Let ${}_{t+1}C^{(i,j)}$ denotes the cash flow at time $\ell + 1$
   if the subject is in state #i and state #j at time $\ell + 1$

   b. Assume the subject is now in state #s at time n

   c. Then the actuarial present value at time n of these cash flows given by the triple product summation (3πΣ)

$$APV_{s@n}\ (C^{(i)}) = \sum_{k=0}^{\infty} [\ {}_{k}Q_{n}^{(s,i)}]\ [\ {}_{n+k}C^{(i)}\ ]\ [\ {}_{k}v_{n}\ ]$$

E. Benefit premiums and reserves

1. Example: benefit premiums

   a. Determined by the equivalence principle
   b. At the time of policy issue, the actuarial present value of premiums should equal the actuarial present value of benefits

2. Example: benefit reserves

   a. The actuarial present value of future loss
   b. Benefits out less benefit premiums in

LAM-138-19
A PRACTITIONER'S GUIDE TO GENEALIZED LINEAR MODELS

Section 1

I. Background

A. Traditional ratemaking in the US is not statistically sophisticated; claims are analyzed using a one-way or two-way analysis

B. Classic linear models and most common bias procedures are special cases of generalized linear models (GLMs)

C. GLMs
   1. Permits explicit assumptions about insurance data and its relationship with predictive variables
   2. More efficient than iteratively standardized models
   3. Provide statistical diagnoses which aid in selecting significant variables
   4. Widely used in ratemaking and underwriting personal lines (auto) insurance

II. The failings of one way analysis

A. Actuaries have relied on one-way analysis for pricing and monitoring performance

B. One-way analysis

   1. Summarizes insurance statistics without accounting for other variables
   2. Explanatory variables can either be discrete (factors) or continuous (variates)
   3. Distorted by correlations between rating factors because factors can be double counted; distortions are resolved today by standardizing the effects of one or more factors
   4. Interdependencies are not considered

C. Multivariate analysis adjusts for correlations and allow investigation into interactions

III. The failings of minimum bias procedures

A. Minimum bias procedures impose a set of equations relating to the observed data, rating variables and a set of parameters that are applied iteratively to converge to a solution

B. The method
   1. Gives no systematic way of testing whether a variable influences the result with statistical significance
   2. Provides no credible range for parameter estimates
   3. Lack a framework to assess quality of modeling

IV. The connection of minimum bias to GLM

A. Common bias procedures

	Generalized Linear Model	
Minimum Bias Procedure	**Link Function**	**Error Function**
Multiplicative balance principle	Logarithmic	Poisson
Additive balance principle	Identity	Normal
Multiplicative least squares	Logarithmic	Normal
Multiplicative MLE with exponential density function	Logarithmic	Gamma
Multiplicative MLE with normal density function	Logarithmic	Normal
Additive MLE with normal density function	Identity	Normal

MLE is the maximum likelihood estimator

B. Not all minimum bias procedures have a GLM model analog

V. Linear models

A. Introduction to linear models

1. A GLM is a generalized form of a linear model

2. Linear models (LMs) and generalized linear models express the relationship between an observed response variable Y and a number of predictor variables (covariates) X

3. Y_i are observations of Y

4. Y is the sum of its mean μ and an error term ε, $Y = \mu + \varepsilon$

5. It assumes that

   a. The expected value of Y, μ, can be written as a combination of the covariates X
   b. The error term is normal with mean 0 and variance σ^2

B. Example

1. Assuming 4 covariates (4 rating factors), a linear model could be

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

2. However there is a linear dependency between the covariates, which means the model is not uniquely defined

3. To uniquely define the model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

4. It assumes that two rating factors are used and the third is a constant added to each

5. Results in the series of equations (assuming X_i has a value of 0 or 1)

$$Y_1 = \beta_1 + 0 + \beta_3 + \varepsilon$$
$$Y_1 = \beta_1 + 0 + 0 + \varepsilon$$
$$Y_1 = 0 + \beta_2 + \beta_3 + \varepsilon$$
$$Y_1 = 0 + \beta_2 + 0 + \varepsilon$$

6. Solve for β by minimizing the sum of the squared errors by taking partial derivatives with respect to β_1 , β_2 , β_3 and setting each to zero

C. Vector and matrix notation

a. Vector notation is used to express these equations into compact form

$$\underline{Y} = \beta_1 \underline{X_1} + \beta_2 \underline{X_2} + \beta_3 \underline{X_3} + \underline{\varepsilon}$$

b. The system of equations takes the form

$$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$$

c. If there are n observations and p parameters in the model, $\underline{\varepsilon}$ will have n components and $\underline{\beta}$ will have p components ($n > p$)

D. Basic ingredients for a linear model consist of

1. A set of assumptions about the relationship between $\underline{Y}$ and its predictor variables

2. An objective function that is optimized to solve the problem

a. Statistical theory defines the likelihood function as a maximum likelihood function

b. Assuming normal error, the parameters that minimize the sum of the squared error also maximize likelihood

VI. Classical linear and model assumptions

A. Linear models

1. Assume all observations are independent and come from a normal distribution

2. The mean is a linear combination of the covariates and each component of the random variable has a common variance

3. The linear model is

$$\underline{Y} = \mathrm{E}[\underline{Y}] + \underline{\varepsilon}$$

$$\mathrm{E}[\underline{Y}] = \boldsymbol{X}.\underline{\beta}$$

4. The explicit assumptions are

a. (LM1) Random component

i. Each component of $\underline{Y}$ is independent and normally distributed
ii. The mean μ of each component is allowed to differ, but all have common variance σ^2

b. (LM2) Systematic component—the p covariates ae combined to give the linear predictor

$$\eta = \boldsymbol{X}.\underline{\beta}$$

c. (LM3) Link Function—specifies the relationship between the random variable and systematic components

$$\mathrm{E}[\underline{Y}] \equiv \mu = \eta$$

B. Limitations of linear models

1. Difficult to assert normality and constant variance for response variables

2. The values of the response variables may be restricted to positive values, violating the assumption of normality

3. If the response variable is strictly non-negative, the variance becomes a function of the mean

4. The additivity of effects in assumptions LM2 and LM3 is not realistic in many applications

VII. Generalized linear model assumptions

A. Assumptions of normality, constant variance and additivity are removed

B. The response variable is now considered to belong to the exponential family of distributions; variance can vary with the mean and the additivity of covariates is assumed on a transformed scale

C. The explicit assumptions are

1. (GLM1) Random component--Each component of $\underline{Y}$ is independent and comes from the exponential family of distributions

2. (GLM2) Systematic component—the p covariates are combined to give the linear predictor

$$\eta = \boldsymbol{X}.\underline{\beta}$$

3. (GLM3) Link Function—specifies the relationship between the random variable and systematic components

$$E[\underline{Y}] \equiv \mu = g^{-1}(\eta)$$

D. Exponential family of distributions

1. The exponential family is a two parameter family defined as

$$f_i\,(y_i;\theta_i;\,\phi) = \exp[((y_i\theta_i - b\theta_i\,)\ \ /\ a_i(\phi))\quad + c\,(y_i\,,\ \phi)]$$

where

$a_i(\phi)$), $b\theta_i$ and $c\,(y_i\,,\ \phi)$ are functions specified in advance

θ_I is a parameter related to the mean

ϕ is the scale parameter related to the variance

2. The exponential family has two properties

a. The distribution is completely specified in terms of mean and variance
b. The variance of Y_i is a function of its mean

3. The second property is emphasized when the variance is expressed as

$$\text{Var}[Y_i] = (\phi\ V(\mu_i))\ /\ \omega_i$$

4. The following distributions belong to the exponential family

Distribution	***V*(*x*)**
Normal	1
Poisson	x
Gamma	x^2
Binomial	$x(1 - x)$ where the number of trials is 1
Inverse Gaussian	x^3

5. Tweedie distribution

   a. Has a point mass at zero and a variance proportional to μ^p where $p < 0$, $1 < p < 2$, *or* $p > 2$

   b. Used to model pure premium data

6. The choice of variance function affects the GLM results

   a. Normal variance assumes each observation has the same fixed variance with equal weight

   b. Poisson variance assumes the variance increases with the expected value of each observation

   c. The gamma variance is more strongly influenced by the point to the left than the point to the right

   d. In addition to $V(x)$, two parameters define the variance of each observation, the scale parameter ϕ and the prior weights ω_i

$$\text{Var}[Y_i] = (\phi\, V(\mu_i)) \,/\, \omega_i$$

E. Prior weights

1. Prior weights allow information about the known credibility of each observation to be incorporated in the model

2. Can be done by weighting exposures; higher exposures will have a lower variance

3. Claim frequency

   a. Define

      m_{ik} The number of claims arising from the k^{th} unit of exposure in cell i

      ω_I The number of exposures in cell i

      Y_i The observed claim frequency in cell i

b. Then

$$Y_i = (1 / \omega_i) \sum_{k=1}^{\omega i} m_{ik}$$

c. If the random process generating m_{ik} is Poisson with frequency of f_i then

$$E[m_{ik}] = f_i = \text{Var}[m_{ik}]$$

d. If exposures are independent,

$$\mu_I = f_i$$

$$\text{Var}[Y_i] = \mu_i / \omega_i$$

e. Since $V(\mu_i) = \mu_I$, $\phi = 1$, and prior weights are the exposures in cell i

4. Claim severity

a. Define

z_{ik} The claim size of the k^{th} claim in cell i

ω_I The number of claims in cell i

Y_i The observed mean claim size in cell i

b. Then

$$Y_i = (1 / \omega_i) \sum_{k=1}^{\omega i} z_{ik}$$

c. Assuming the random process is random

$$E[z_{ik}] = m_i$$

$$\text{Var}[z_{ik}] = \sigma^2 m_i^2$$

d. And if each claim is independent,

$$\mu_I = m_i$$

$$\text{Var}[Y_i] = \mu_i^2 \sigma^2 / \omega_i$$

e. Since severity is a gamma distribution, $V(\mu_i) = \mu_i^2$, $\phi = \sigma^2$, and prior weights are the number of claims in cell i

F. The scale parameter

1. Under the Poisson distribution, the scale parameter equals 1; in other cases it must be estimated from the data

2. It is not needed to solve the GLM but is needed for certain statistics, such as ϕ

3. ϕ can be treated as a parameter and estimated by maximum likelihood; the drawback is the inability to derive an explicit formula for ϕ

4. An alternative is to use an estimate of ϕ

   a. The moment estimator (Pearson χ^2 statistic)

$$\widehat{\emptyset} = \frac{1}{n-p} \sum_i \left[(\omega_i (Y_i - \mu_i)^2) / V(\mu_i) \right]$$

   b. The total deviance estimator

$$\widehat{\emptyset} = \frac{D}{n-p} \text{ where } D \text{ is the total variance}$$

G. Link functions

1. Classic linear regressions require data transformation to satisfy the requirements of normality and constant variance and additivity effects

2. GLM only requires additivity—the transformation of Y must be additive in the covariates

3. Consider μ_i as the function of the linear predictor $\mu_i = g^{-1}(\eta)$

4. The link function must be monotonic and differential

5. Typical choices for a link function are:

Link Function	$g(x)$	$g^{-1}(x)$
Identity	x	X
Log	$\ln(x)$	e^x
Logit	$\ln(x/(1-x))$	$e^x / (1 + e^x)$
Reciprocal	$1 / x$	$1 / x$

6. Each error structure has a link function that simplifies the math

7. The log link function property is that the effect of the covariates are multiplicative

8. When a log link function is used, the GLM estimates logs of multiplicative effects

H. The offset term

1. When the effect of the explanatory variable is known, rather than estimating $\underline{\beta}$ in respect of this variable, it is appropriate to include the known effect of this variable

2. This is done by adding an offset term $\underline{\xi}$ into the definition of linear predictor η

$$\eta = \boldsymbol{X}.\underline{\beta} + \underline{\xi}$$

$$E[\underline{Y}] \equiv \mu = g^{-1}(\eta) = g^{-1}(\boldsymbol{X}.\underline{\beta} + \underline{\xi})$$

3. If the expected count of claims increases in proportion to the exposure of an observation, it can be introduced in a multiplicative GLM by setting the offset term to equal the log of the exposure of each observation

4. A Poisson multiplicative GLM modeling claim counts with an offset term equal to the log of the exposure produces identical results to modeling claim frequencies with no offset term but with prior weights equal to the exposure of each observation

I. Structure of a generalized linear model

$$\mu_I = E[Y_i] = g^{-1}\left(\sum_{\boldsymbol{j}} X_{ij}\beta_j + \xi_i\right)$$

$$\text{Var}[Y_i] = (\phi\, V(\mu_i)) \,/\, \omega_i$$

where

Y_i is the vector of responses

X_{ij} is a matrix (the design matrix) produced from the factors

$g(x)$ is the link function: a specified function that relates the expected response to the linear combination of observed factors

β_j is the vector of model parameters to be estimated

ξ_I is the vector of known effects or offsets

ϕ is a parameter to scale $V(x)$

$V(x)$ is the variance function

ω_i is the prior weight that assigns a credibility or weight to each observation

VIII. Typical GLM model forms

A. Typical model is multiplicative Poisson

B. It is intuitively appropriate because it is invariant to measures of time (not true for distributions such as gamma)

C. Weights are set to be the exposure of each record; for claim counts, it is the log of the exposure

D. A common model form is multiplicative gamma; it has an intuitively attractive property for modeling claim amounts because it is invariant to currencies (not true for distributions such as Poisson)

E. The typical model form for modeling retention and new business is a logit link function and binomial error term; the logit function is invariant to measuring success or failure and it can be approximated by a Poisson function if used qualitatively and the *y*-variates are generally close to 0

F. Typical model forms are summarized below

$\underline{Y}$	Claim Frequencies	Claim Numbers or Counts	Average Claim Amounts	Probability
Link function $g(x)$	ln(x)	ln(x)	ln(x)	ln(x/(1-x))
Error	Poisson	Poisson	Gamma	Binomial
Scale parameter ϕ	1	1	Estimated	1
Variance function $V(x)$	x	x	x^2	x(1-x)
Prior weights $\underline{\omega}$	Exposure	1	# of claims	1
Offset $\underline{\xi}$	0	ln(exposure0	0	0

IX. Maximum likelihood estimators

A. After defining a model in terms of $\boldsymbol{X}$, $g(x)$, ξ_I, $V(x)$, ϕ and $\underline{\omega}$ and a set of observations $\underline{Y}$, the components of $\underline{\beta}$ are derived by maximizing the likelihood function (or equivalently, the logarithm of the likelihood function)

B. Method finds parameters that produce the observed data with highest probability

C. The likelihood is defined as the product of probabilities of observing each value of the *y*-variate

1. For continuous distributions, the probability density function is used

2. The log of the likelihood is used because it makes calculations more manageable

3. In simple examples, the likelihood is found by finding a solution to a system of equations; for complex examples, numerical methods are used

X. General procedure for solving a GLM involves the following steps

A. Specify the design matrix $\boldsymbol{X}$ and the vector of parameters $\underline{\beta}$
B. Choose the error structure and link function
C. Identify the log-likelihood function
D. Take the logarithm to convert the product of may terms into a sum
E. Maximize the logarithm of the likelihood function by taking partial derivatives with respect to each parameter, setting them to 0 and solving the resulting system of equations
F. Compute the predicted values

XI. Solving for large datasets using numerical techniques

A. Iterative numerical methods are used

1. Optimize the likelihood of the values of $\underline{\beta}$ which set the first differential of the log-likelihood to zero

2. Newton Raphson iteration process can be used

$$\underline{\beta}_{n+1} = \underline{\beta}_n - \boldsymbol{H}^{I}.\underline{s} \qquad \text{where}$$

$\underline{\beta}_n$ is the n^{th} iterative estimate of the vector of parameter elements $\underline{\beta}$ (with p elements)

$\underline{S}$ is the vector of first derivatives of the log-likelihood

$\boldsymbol{H}$ is the p by p matrix containing second derivatives of the log-likelihood

3. To start, set parameters to 0 or use results from a one-way analysis

B. Base levels and intercept term

1. In a three factor model, the overall average response was incorporated into two other rating variables; the decision as to which variables are used is arbitrary

2. When considering many factors, parameterize GLM by considering an intercept term that applies to all observations

3. When considering categorical factors and an intercept term, one level of each factor should have no parameter associated with it, in order that the model remains uniquely defined

4. If a model is structured with an intercept term but without each factor having a base level, the GLM routine removes as many parameters as necessary to uniquely define the model (known as aliasing)