Predictive Modeling for Smaller Companies – Commercial

CAS Special Interest Seminar on Predictive Modeling

Richard A. Smith, FCAS, MAAA

Katharine Barnes, FCAS, MAAA

October 4, 2006

Predictive Modeling in Commercial Lines

- Overall rates for smaller companies writing commercial lines

    - Tend to be set with Bureau experience

    - Modified by company experience

    - Further modified by competitive landscape

- Market is becoming more complex

    - Many larger carriers are applying approaches they've learned from personal lines

- Predictive modeling can be used to refine rate levels

    - Tiers/underwriting (who is really "preferred")

    - Schedule rating (more data-driven than underwriters' judgment)

    - Others

What is a Smaller Company to Do?

- If we do nothing, adverse selection will happen.

- But we don't have enough data to do this type of analysis…

A Smaller Company CANNOT Just…

- Add on "credit score" based on a competitor's filing, on top of the existing rating plan

    - This over-discounts some classes, and under-discounts others

    - May have implications for new business growth

- Create a full-blown class plan on their own data

A Smaller Company CAN Do Something

- For predictive modeling projects, a range of information can be used to analyze the relative rate levels for different types of insureds

- The more data you have, the more complex an analysis can be performed, but you do not need to be a "Top 10" carrier to have enough data to do some level of analysis.

    - Large volume of very detailed company-specific data

        — Rate plans, rate relativities, tiers, territories

    - Smaller volume of data (fewer records) and/or fewer variables captured per record suggests scaled back analyses

        — Tiers and tier relativities

        — Refine major risk factors (subdivide categories)

        — Underwriting and/or schedule debit/credit guidance

A Smaller Company May Have Some Advantages

- You know your local market

- You know your agents

  - May be a good source of competitive information

- You may be able to implement changes without drawing too much attention from your competitors

- You may be able to "Me Too" parts of some competitors' filings

- You may have a stable history

  - Fewer mergers and acquisitions – fewer legacy systems issues

  - More institutional knowledge

  - Better communication within the company (underwriting, claims, marketing can all talk to actuarial)

- More streamlined management decision-making

How Much Data is Enough?

- Actuaries will always welcome more data but…

  - Think claims, not exposures

  - At a minimum 5,000 claims are needed

    — This is a VERY ROUGH rule of thumb

    — Much depends on what you're trying to analyze

  - For a relatively high frequency line of business, fewer exposures are needed

    — Can combine information – for example for commercial automobile, can combine coverages for relativities if the volume is too low for each coverage individually (e.g., BI/PD)

    — Can add more years worth of data

  - A model of frequency (rather than severity or pure premium) can fit the data more easily

How Much Data is Enough?

- Hypothetical Examples

  - Commercial Auto – 5,000 claims at frequency of 5% = 100,000 exposures, or 25,000 exposures per year over four years

    — Significantly lower frequency for BI – would need adjustment to analysis

  - BOP – 5,000 claims at frequency of 10% = 50,000 exposures, or 12,500 exposures per year over four years

    — Issues related to liability versus property coverages

  - Workers Compensation

    — Similar concepts

Types of Data

■ If you do not have data

   ■ You can buy it

   ■ You can find it (including in your own underwriting department)

   ■ You can start collecting it now, and use it later

   ■ You can review relativities from competitors to estimate the effect of the data

Types of Data for Commercial Automobile

- Personal Auto

  - Driver characteristics

  - Vehicle characteristics

  - Household characteristics

  - Other insurance characteristics

  - Agent characteristics

  - Etc.

  - (Hundreds of possibilities)

- Commercial Auto

  - In many cases a similar or parallel type of data element is useful, especially for smaller commercial auto policies (single vehicles or small fleets)

    — For sole proprietors, very similar data

    — For small insureds, driver characteristics may be available, vehicle data will be available, and other characteristics may be purchased

Types of Data for Commercial Automobile

- Vehicle Information

  - Number, type, age, weight, value, location, etc.

- Driver Information

  - Number, age, sex, marital status, credit, driving record, etc.

- Business

  - Years in business, size of business, class of business, etc.

- Owners

- Environmental/economic

- Claims history

  - For commercial auto and other coverages

- Policy

  - Duration, deductibles, endorsements, other coverages

- Agent

Types of Data for Business Owners Policies

- Building Information

  - Construction class, age, amount of insurance, location, etc.

- Employee Information

  - Number, demographics, classes, etc.

- Business

  - Years in business, size of business, class of business, etc.

- Owners

- Environmental/economic

- Claims history

  - For BOP and other coverages

- Policy

  - Duration, deductibles, endorsements, other coverages

- Agent

Types of Data for Workers Compensation

- Company Information

  - Years in business, size of business, class of business, etc.

- Employee Information

  - Number, demographics, classes, etc.

- Safety Programs

  - Implementation, effectiveness, OSHA violations, etc.

- Owners

- Environmental/economic

- Claims history

  - For WC and other coverages

- Policy

  - Duration, deductibles, endorsements, other coverages

- Agent

Types of Data for Medical Malpractice

- Medical Practice Information

  - Specialty, years in business, size of business, stability of members, etc.

- Employee Information

  - Number, demographics, classes, etc.

- Owners

- Environmental/economic

- Claims history

  - For medical malpractice and other coverages

- Policy

  - Duration, deductibles, endorsements, other coverages

- Agent

External Data

- You can buy data, or get it from publicly available sources

  - Insured-Specific

    — Credit score

    — Corporate financial information

    — Insured reported infractions

  - Location-Specific

    — Census/demographics

    — Weather

    — Crime statistics

- Data may not be available for all business, or may not be available in all states

  - Quality of data can vary

Data Preparation is 80% of the Battle

- Messy data causes some problems
  - Model may not converge (you'll have to fix the data anyway!)
  - Poor coding will suppress differences between classes

- Hypothetical example - 20% of drivers not coded as youthful (based on statistical system issue, not actual rating)

- Particularly likely in commercial auto

Correct Driver Information

Incorrect Driver Information
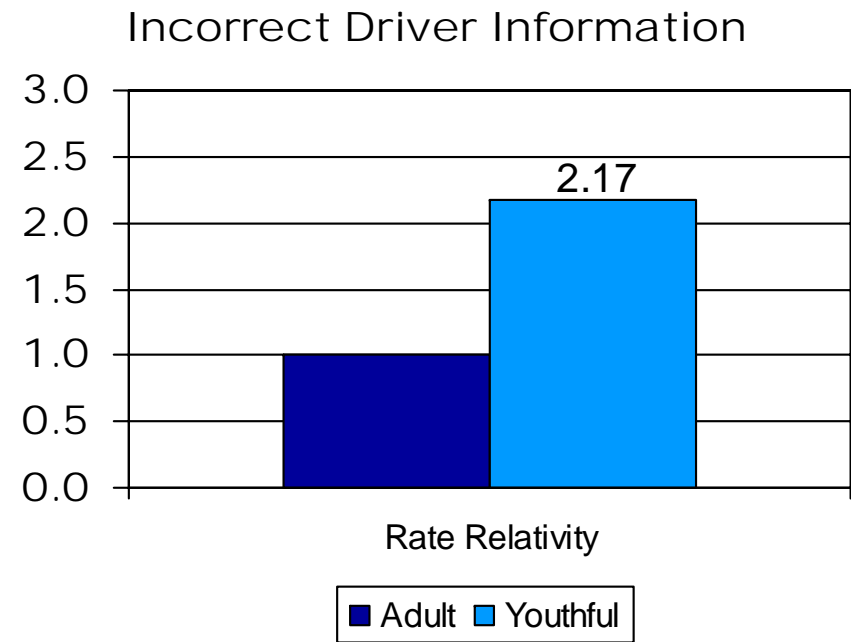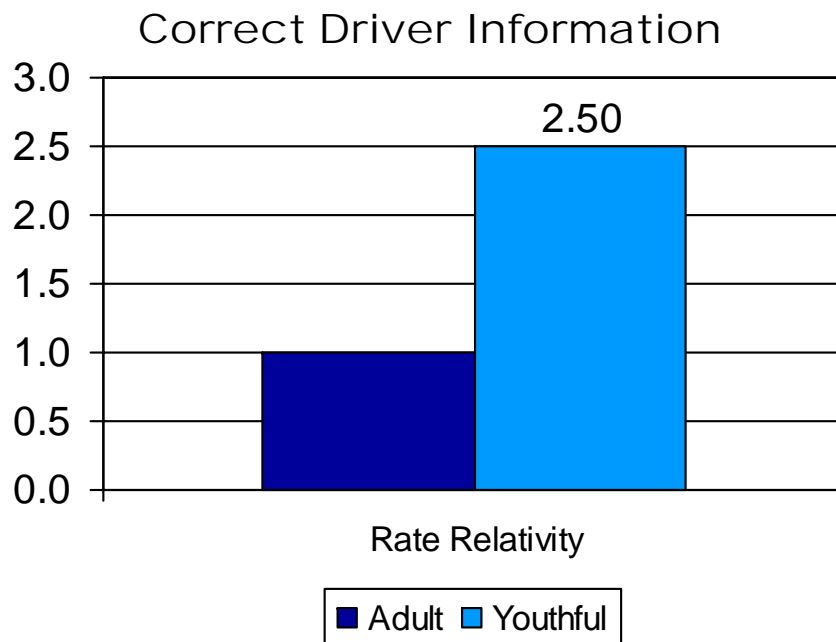
Now You're Ready to Start Analyzing

- Get ready to deal with some pragmatic issues

- Be aware of some modeling issues

  - Potential strategies

  - Sometimes involves trade-offs

Pragmatic Issues

- Understand your goals

  - Underwriting model or tiers probably more forgiving

    — Want to control for rating characteristics, not set those relativities

    — Initially, want direction and magnitude, not precision

  - Subdividing a broad category – similar issues

  - These decisions also drive your data needs

- Variable selection (initial candidates and later, what to leave in model)

  - Brute force searches using predictive modeling is not a substitute for subject matter expertise

  - Speak with underwriters and claim adjusters

    — They'll possibly have a sense for important variables

    — They may have an idea of magnitude

  - Understand competitors' approaches (don't reinvent the wheel)

Modeling Issues

- **Big danger is over-fitting**

  - Find signal not noise

  - Judgment will often be needed in face of volatility

  - When possible split the data (out of sample validation)

- **General understanding of modeling will help maximize value of your data**

  - Data volume

  - Model design

Dealing with Volatility

- Use more years of data

- Look at a given factor's interaction with year

  - A variable that is stable from year to year is more credible, even with low volume of claims

- Cross-validation or re-sampling approaches

  - Create model with most of data, and then test it on the remaining data

- May want to "fix" some factors at predetermined levels

  - Major rating factors

  - Desirable discounts

  - Then let model compensate for remaining variables

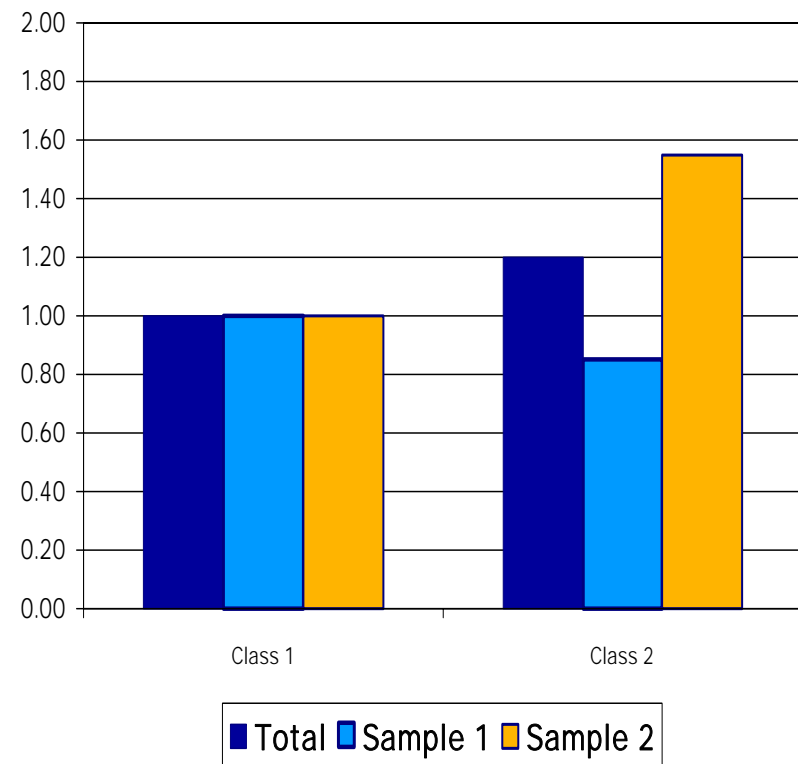# Volatility versus Consistency by Year

Desirable
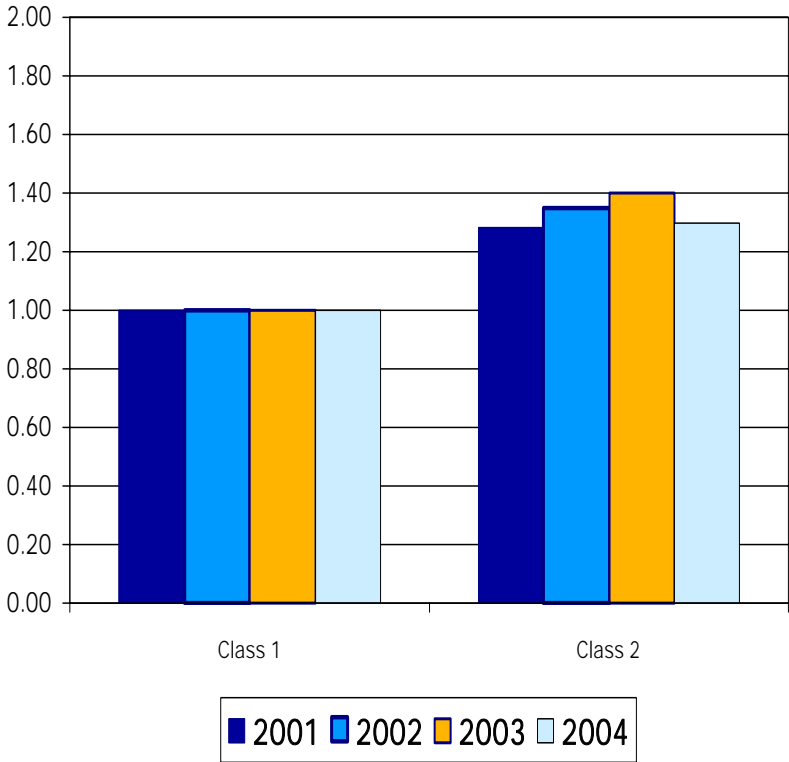
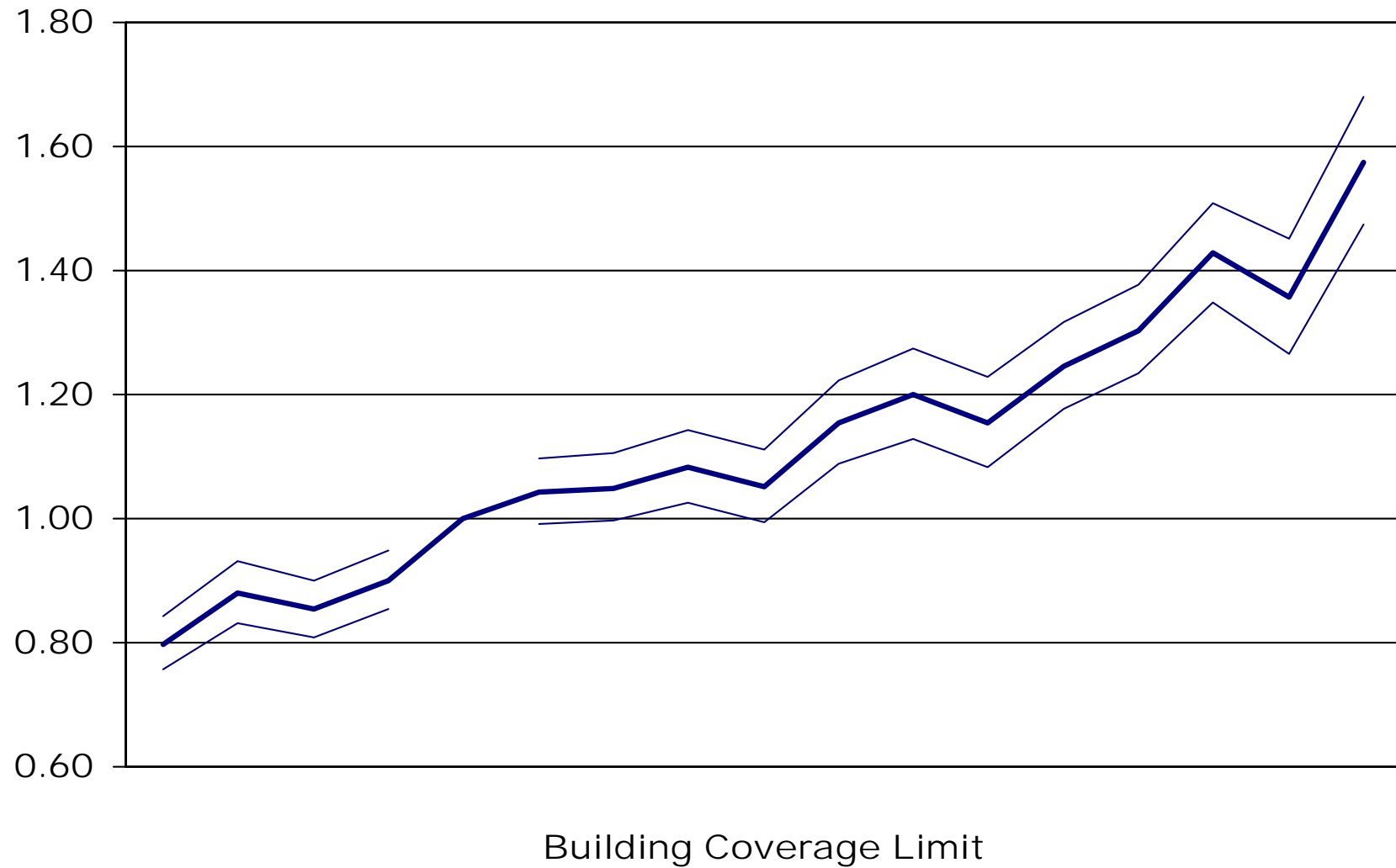Not So Desirable
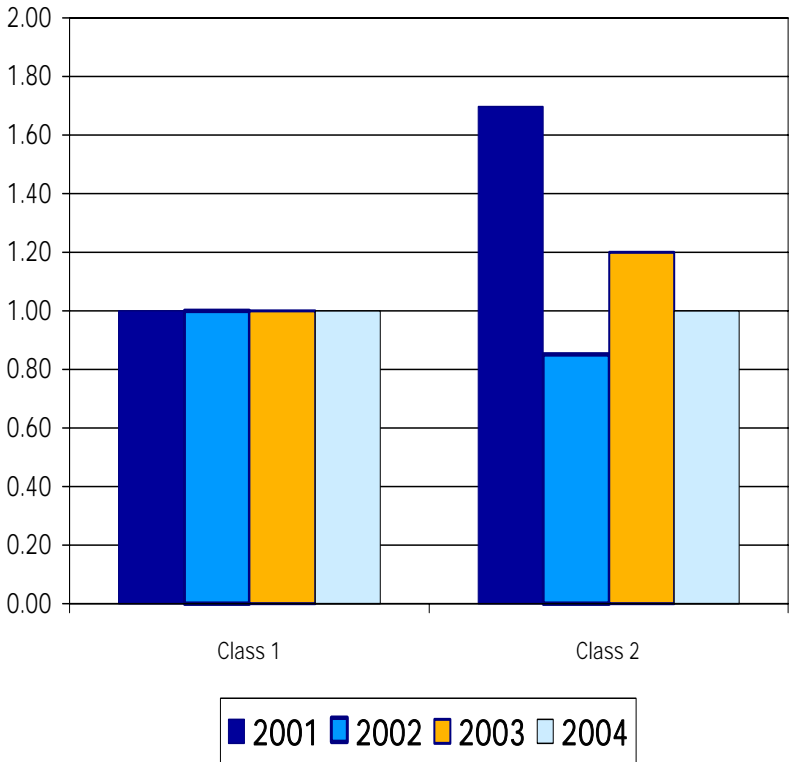
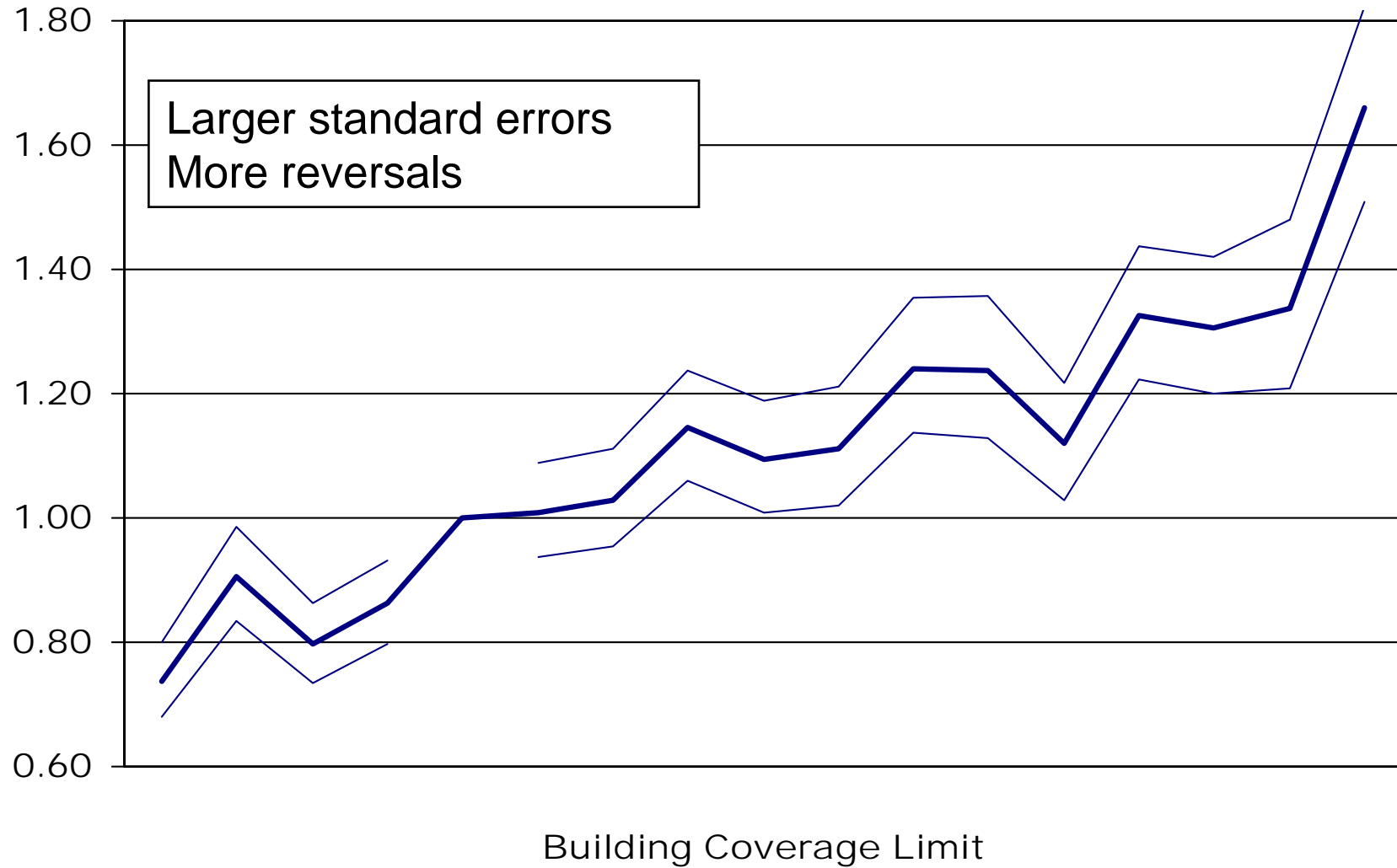# Volatility versus Consistency in Samples
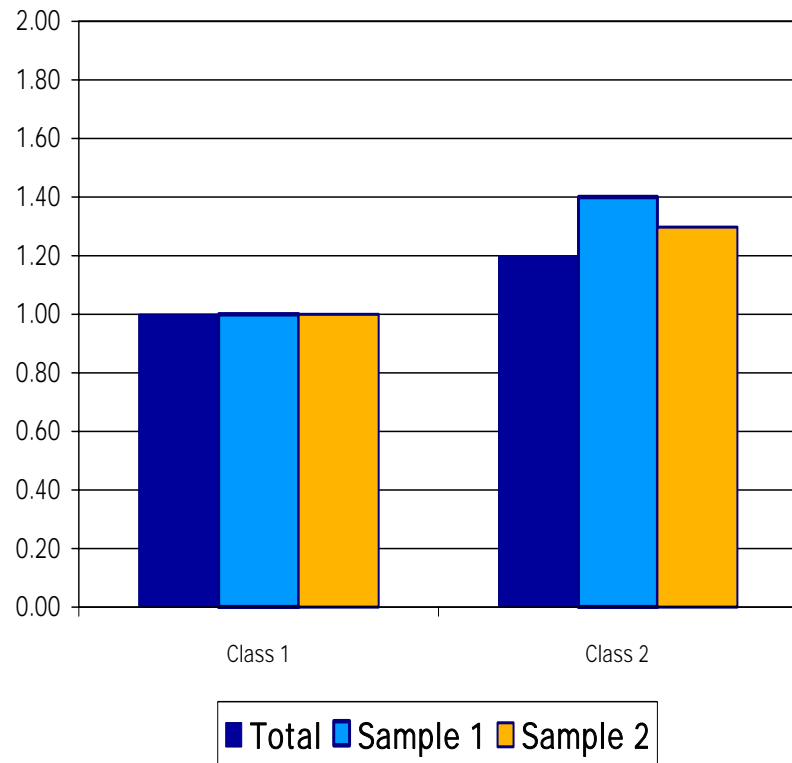
Model Design

- Although it might seem counter-intuitive, consider separate frequency and severity models

    - Frequency often predominate contributor to cost differences

    - Standard errors usually tighter – more variables survive vetting

    - Isolates claim size volatility (often end up with simpler severity models)
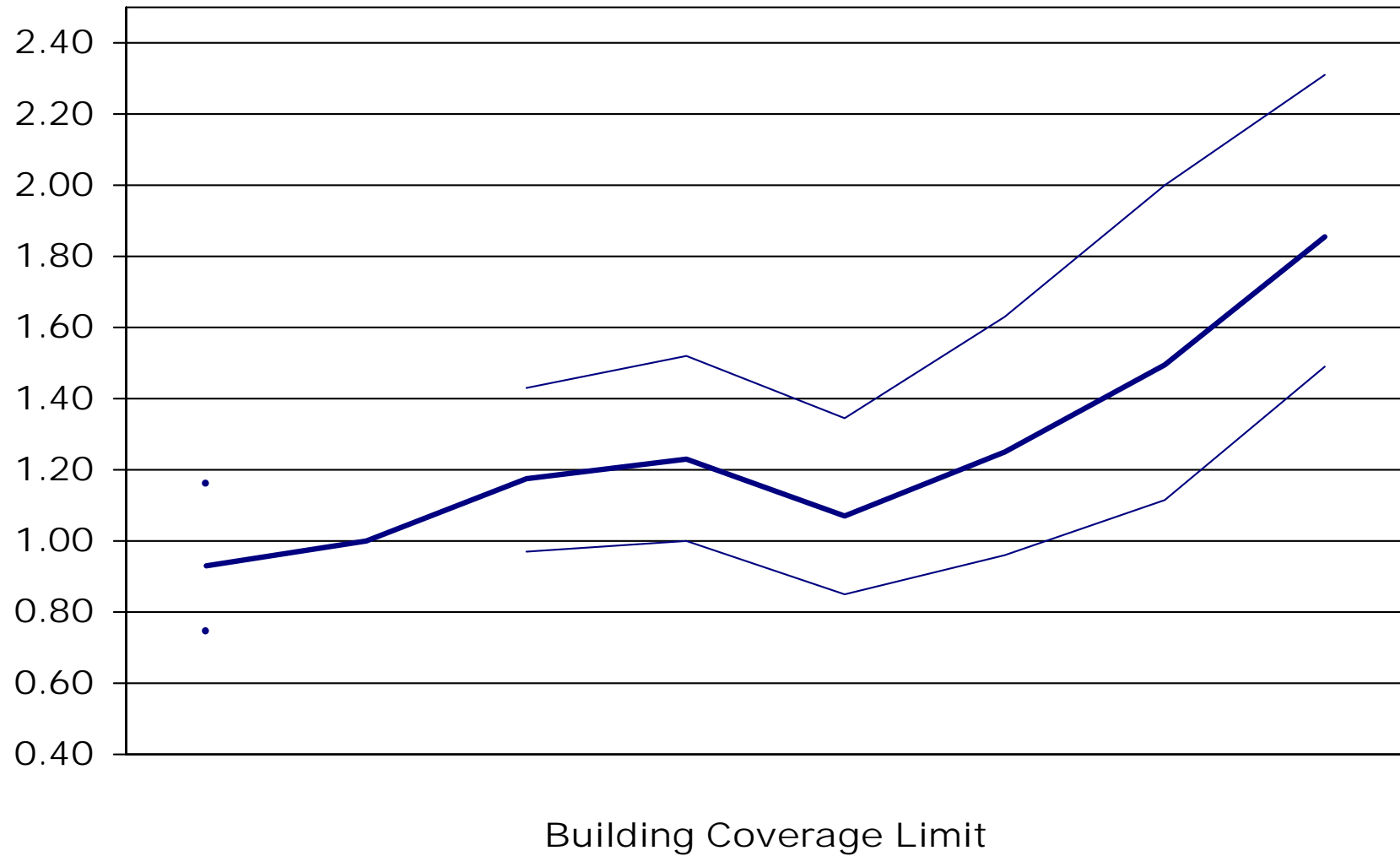
    - Then combine into pure premium model

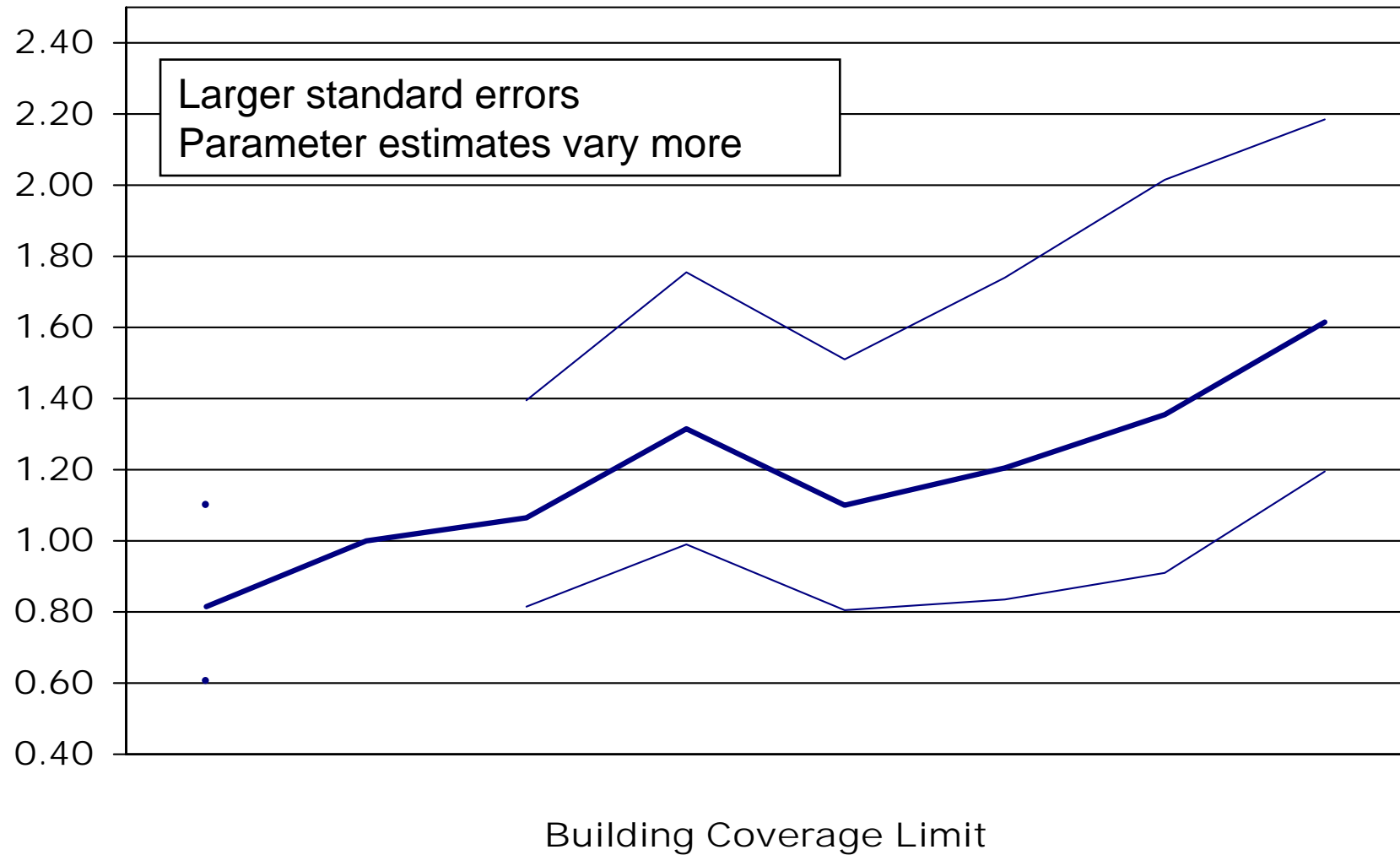# Indicated Frequency Relativities – 10,000 Claims

Building Coverage Limit

Indicated Frequency Relativities – 5,000 Claims

Larger standard errors
More reversals

Building Coverage Limit

Indicated Severity Relativities – 10,000 Claims

Building Coverage Limit

Indicated Severity Relativities – 5,000 Claims

Larger standard errors
Parameter estimates vary more

Building Coverage Limit
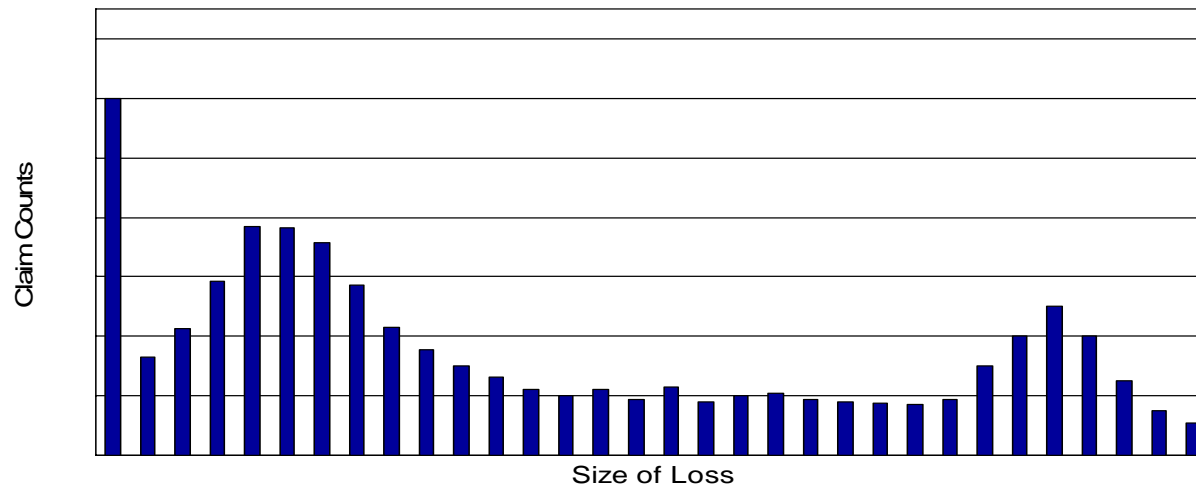
Model Design (Continued)

- **Watch degrees of freedom**

    - Too many variables in a model, or too many levels of the variables

    - Might need to treat some variables as continuous

    - May want to look at piece-wise/spline approaches

Building Coverage Limit

— 5,000 claims — 10,000 claims — Piece-wise
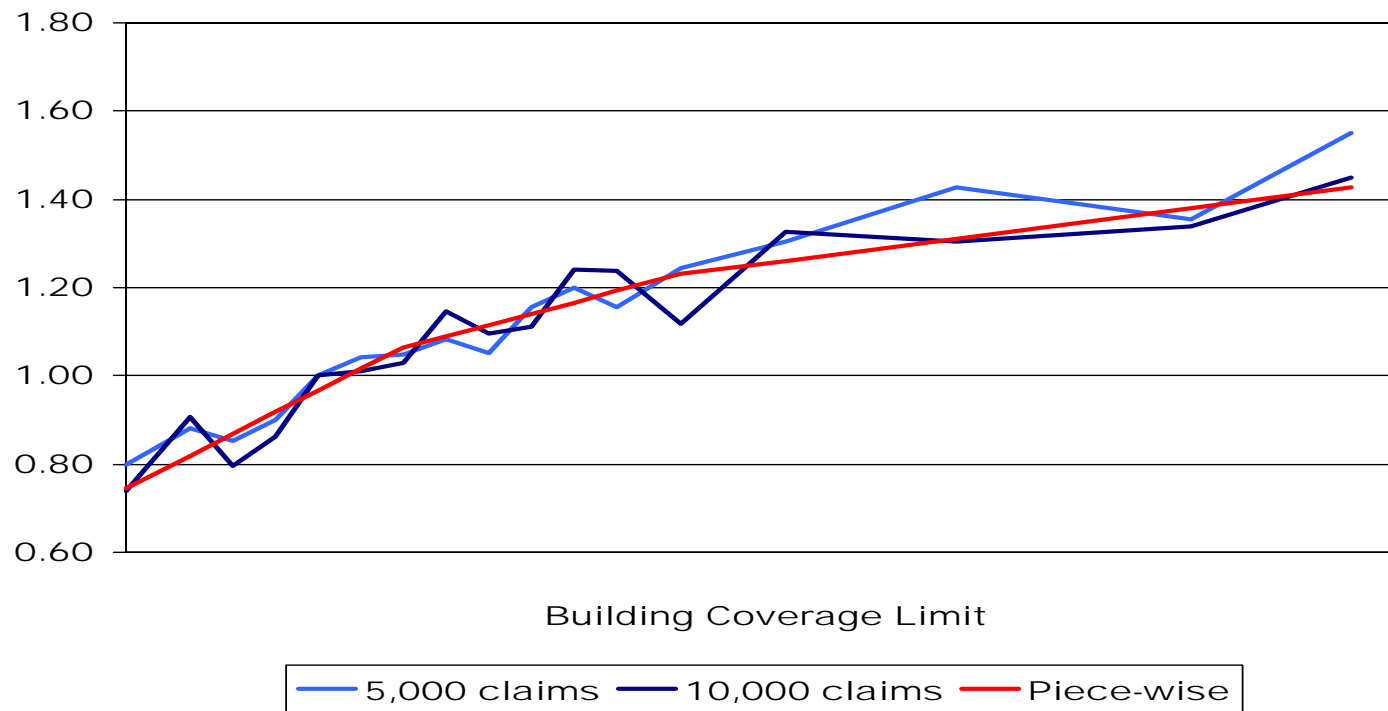
Model Design (Continued)

- Similarly, you may wish to split losses into multiple parts

  - By peril/cause of loss

  - By size of loss

- Should give more homogeneous buckets to work with

- Even so, you still may end up with simple models

Comprehensive Coverage - Hypothetical Data

Model Design (Continued)

- Alternate approach might be able to analyze loss ratios

    - More noise to deal with

    - Need premiums at current rate level

        — Can be difficult to determine

    - Main benefit is in combining coverages

    - Probably good for tier or underwriting model

        — Guidance for schedule debits and credits

Once You Have Data, and Analyze It, What are the Other Major Issues?

- Implementation – Once you get your data in place and do your predictive modeling, and you change your rating structure, different insureds will be affected differently

  - Rate Level Effect – What off-balances do you need to get a revenue-neutral plan (or measure the effect on revenues of the new plan?

  - Dislocation – How do your new indicated rates compare to your current rates?

  - Competition – How do your new indicated rates compare to your competitors' rates?

  - Communication – How do you communicate the new plan to your sales force (either internal/captive or independent agents)?

  - Regulation – How do you get your new rating plan approved?

  - Support – How does your systems department support the new plan?