TOWERS PERRIN
TILLINGHAST

Predictive Modeling for Smaller Companies

CAS Annual Meeting -- Chicago

Richard A Smith

November 12, 2007

What are the concerns for a smaller company considering predictive modeling?

"If we do nothing, adverse selection will happen."

"Do we have enough data to do this type of analysis?"

"What kind of analysis can we do with the data we do have?"

Some specific concerns among smaller companies

- Current trends are in the wrong direction
  - Drop off in policies written
  - Low retention on parts of the book
  - Deteriorating loss ratios
- Some evidence of adverse selection
  - Drift toward worse credit-based insurance scores
- Frustration among producers
  - You always used to be competitive for a multi-car, multi-line clean driving record

Some specific concerns about competitors

- Market leaders already have complex rating plans and other innovations
  - Progressive – the standard for a number of years
  - Others have reacted, and in some cases have gone beyond
    - —Allstate (“Your Choice”)
    - —State Farm (Customer Rating Index)
    - —Travelers (Quantum)
- Move toward even more complexity
  - More tiers
  - More interaction-based rating factors

A smaller company may have some advantages over larger companies

- Better local market knowledge
  - Agents may be a good source of information
- Implementation may not draw too much attention
  - Larger competitors may not care
  - Regulatory objections may be less frequent
- May be more nimble, and able to implement changes faster
  - More institutional knowledge — in a smaller group of people
  - Better communication between departments
  - More streamlined management decision-making
  - One wildcard is capacity/capability of policy processing systems

Exploration is fun, but it's better to have a plan . . .

- Identify business goals
  - Rating plans: Better accuracy, new structure, new variables
  - Retention/elasticity: Problem areas, impact of planned changes
- Identify limitations
  - Timeline
  - Budget
  - Data availability
- Develop your plan
  - Broad questions ➔ more leeway on data and analysis
  - Study of all coverages / perils combined or separately?
  - Will a frequency study suffice?
  - Will the same factor apply to all coverages?

A smaller company should probably not just…

- Add a credit score (or some other variable) 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
- Attempt to create a full-blown class plan on their own data
  - Less reliable results if data are too thin
    - Variables may not be significant
    - Levels within variables may be volatile
  - Models can be unstable as variables are added or dropped

There are things that can be done — subject to limitations

- Given many records and data variables
  - Analyze main rating variables and rate relativities
  - Explore new variables or variable interactions
  - Analyze territory boundaries and relativities
    - Be cautious about calculating directly in a GLM
- Fewer records and/or fewer variables suggest simpler analyses
  - Tier definitions and tier relativities
  - Refine major risk factors (subdivide some categories)
  - Underwriting and/or schedule debit/credit guidance
  - Rely more heavily on competitive analysis for other key variables

Don’t forget about retention models

- Identify groups with better / worse than average retention
- Identify specific events which might influence retention
  - Look at policy change like change or add car, change or add drivers, etc.
  - Rate changes
  - Isolate voluntary vs. involuntary attrition
- Effects of competition

Match model complexity to data and goals

- Underwriting models or tier analysis
  - Probably sufficient to control for main rating characteristics, but not necessarily use results to change those factors
  - Initially, want direction and magnitude, not precision
- Retention — you may want to isolate company actions
  - Insurance to value program, changes in some billing options, etc.
    - Again, control for main characteristics
    - Add appropriate indicator variables
- Elasticity/rate impact analysis adds other twists
  - May want to restrict to short intervals
    - Management may want quick feedback: how did rate change affect mono-line vs. multi-line risks?
  - Lots of exogenous variables change quickly – e.g. competitive information

How many records will I need to model loss data?

- Think claims, not exposures
  - For a relatively high frequency line of business, fewer exposures are needed
- Would prefer 5,000+ claims
  - This is a very rough rule of thumb
  - A lot depends on what you're trying to analyze
- If you're short on claims, pursue more years of data

Identifying variables to include

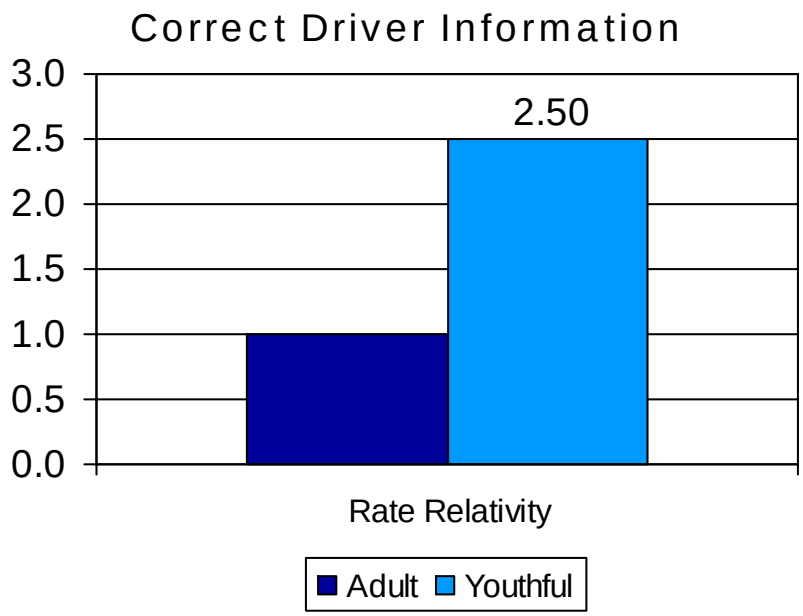
- Brute force searches using predictive modeling is not a good substitute for subject matter expertise
- Speak with underwriters and claim adjusters
  - They'll possibly have a sense for important variables
  - They may have an idea of magnitude
- Understand competitors' approaches (don't reinvent the wheel)

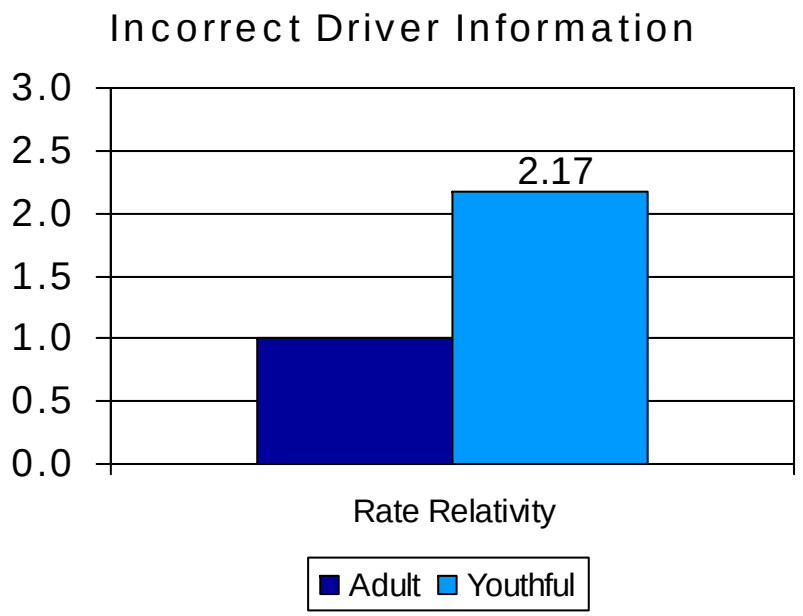
Data preparation is the key to success

- Ideally want near transaction-level data
  - At the policy (vehicle) level
  - A separate record representing the risk characteristics in effect for that period
    - When a variable changes, add a new record
- Prefer to have only one claim per policy exposure segment
  - If more than one claim, can split the premium record into finer segments
  - Can usually aggregate claims to match to a given period
    - Parameter estimates are unchanged
    - Standard errors are understated

Messy data causes problems

- Model may not converge (you'll have to fix the data anyway!)
- Poor coding will suppress differences between classes
- Hypothetical example - 20% of vehicles not coded as youthful
  - Rate relativities are understated for young drivers





Other potential data problems

- Severity analyses – very small claims
  - Make sure they're not miscodes
  - Possibly an issue with salvage/subrogation
  - Saw one instance where LAE was often $5 – for police reports
- Frequency analyses – claims on very small exposure periods
  - Check to see if there are system issues
- Miscoded/missing data
  - Often highly correlated for several independent variables
  - "Near-aliasing" can result, and cause convergence problems

Construction / Fire Protection	1-4	5-8	9	10	Miscoded Fire Protection
Brick/Fire Resistive	3000	2000	800	500	0
All Other Valid Types	6000	4500	1600	1000	5
Miscoded Construction	0	0	0	0	395

  - One option – force "All Other Valid Types"/"Miscoded Fire Protection" to "Miscoded Construction"

Potential ways to simplify data preparation

- Accumulate data on a policy year basis
- Take snapshot at beginning of period
- Take snapshots quarterly
- Worry about “important” policy changes
  - Change in vehicles
  - Change in driver
  - Add/drop coverage
  - Ignore things like change in lien holder

Do you already have what you need?

- If your current data system has:
  - Merged premium and loss data partially summarized
  - Includes all major characteristics
  - Individual claim data is separately available with risk characteristics attached
  - You could:
    - Build a frequency model on aggregated results
    - Do a separate claim severity model
    - Combine frequency and severity factors manually
- If individual claim data are not available
  - You could do pure premium analysis directly using a Tweedie distribution
- It may not be possible to add new variables when using summarized data

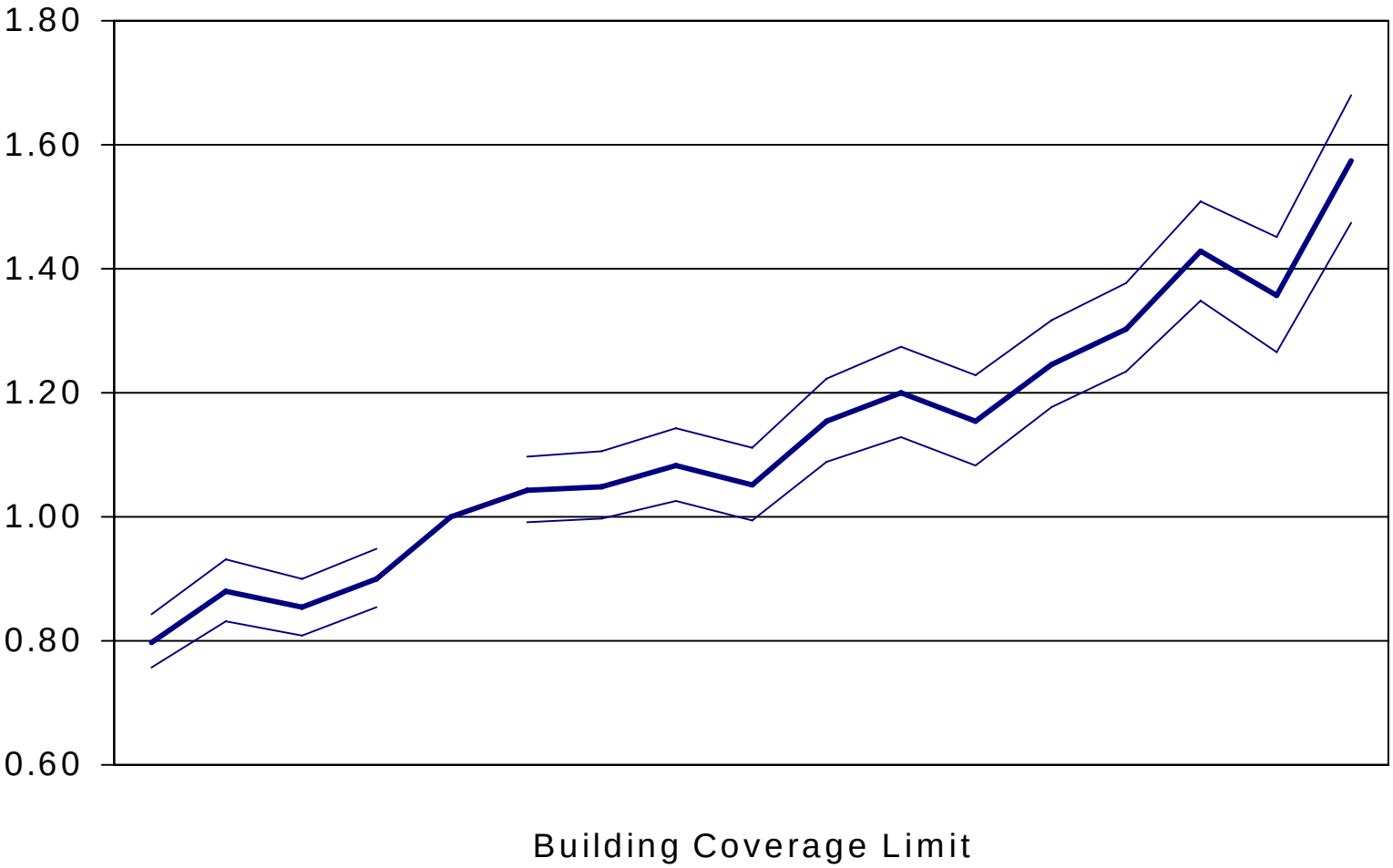
Model design

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

- Frequency is often the predominate contributor to cost differences
- Standard errors are usually tighter – more variables survive vetting
- Isolates claim size volatility (often end up with simpler severity models)
- Then combine into pure premium relativities

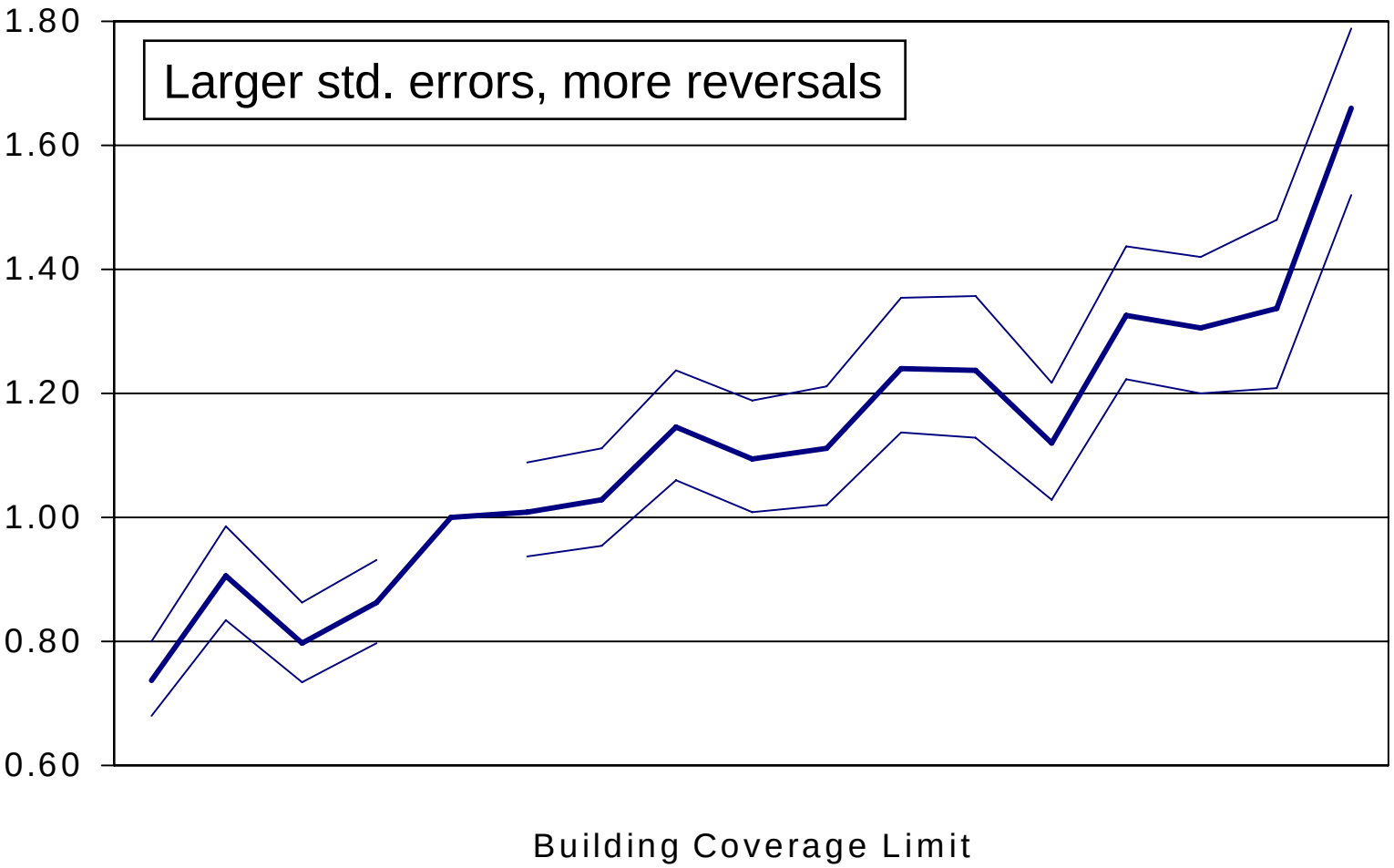
Indicated frequency relativities – 10,000 claims

ILLUSTRATIVE



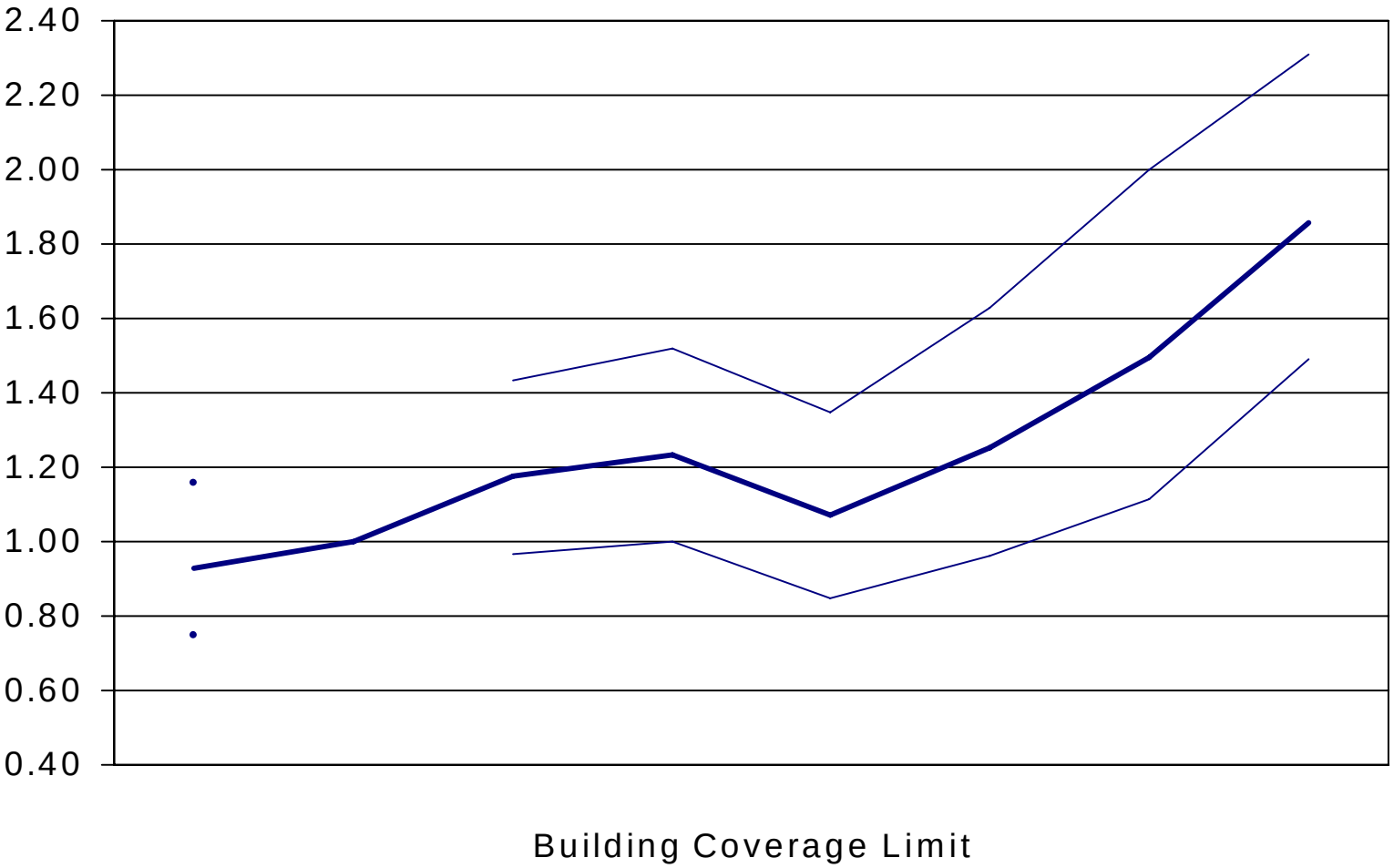
Indicated frequency relativities – 5,000 claims

ILLUSTRATIVE



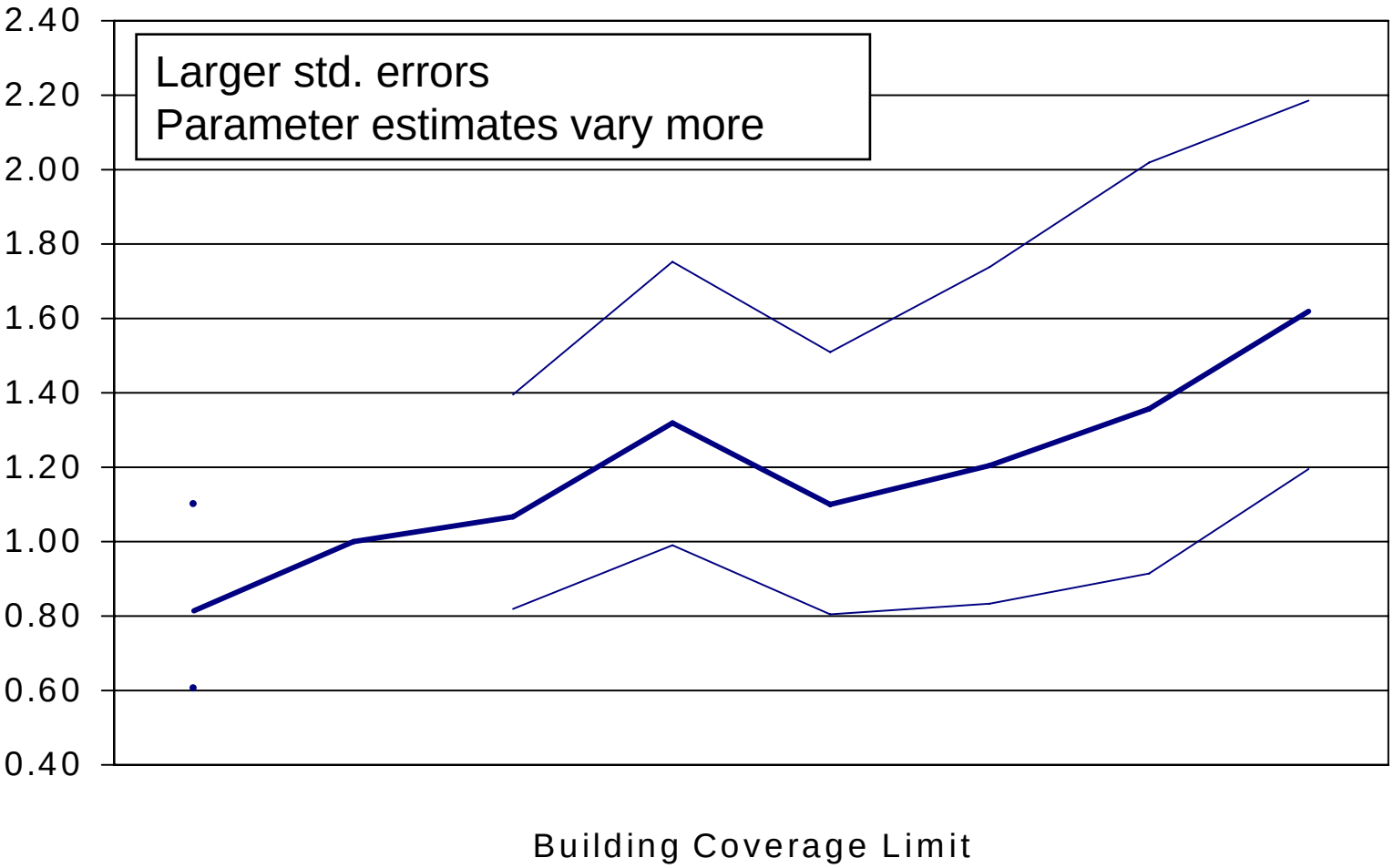
Indicated severity relativities – 10,000 claims

ILLUSTRATIVE



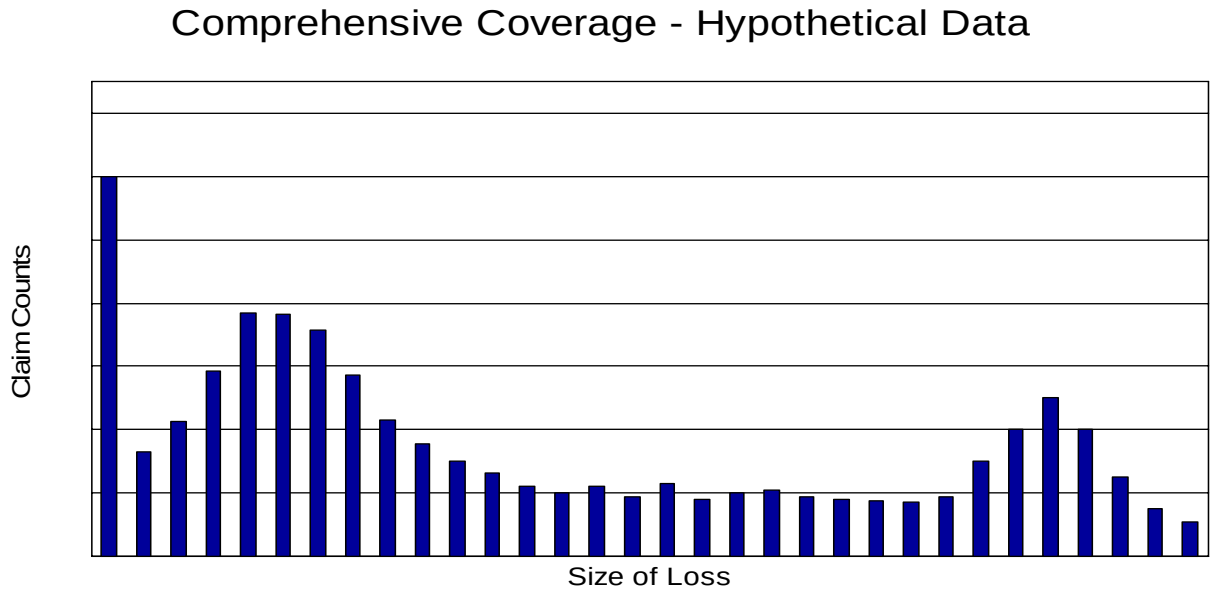
Indicated severity relativities – 5,000 claims

ILLUSTRATIVE



Effect of claim type on model design

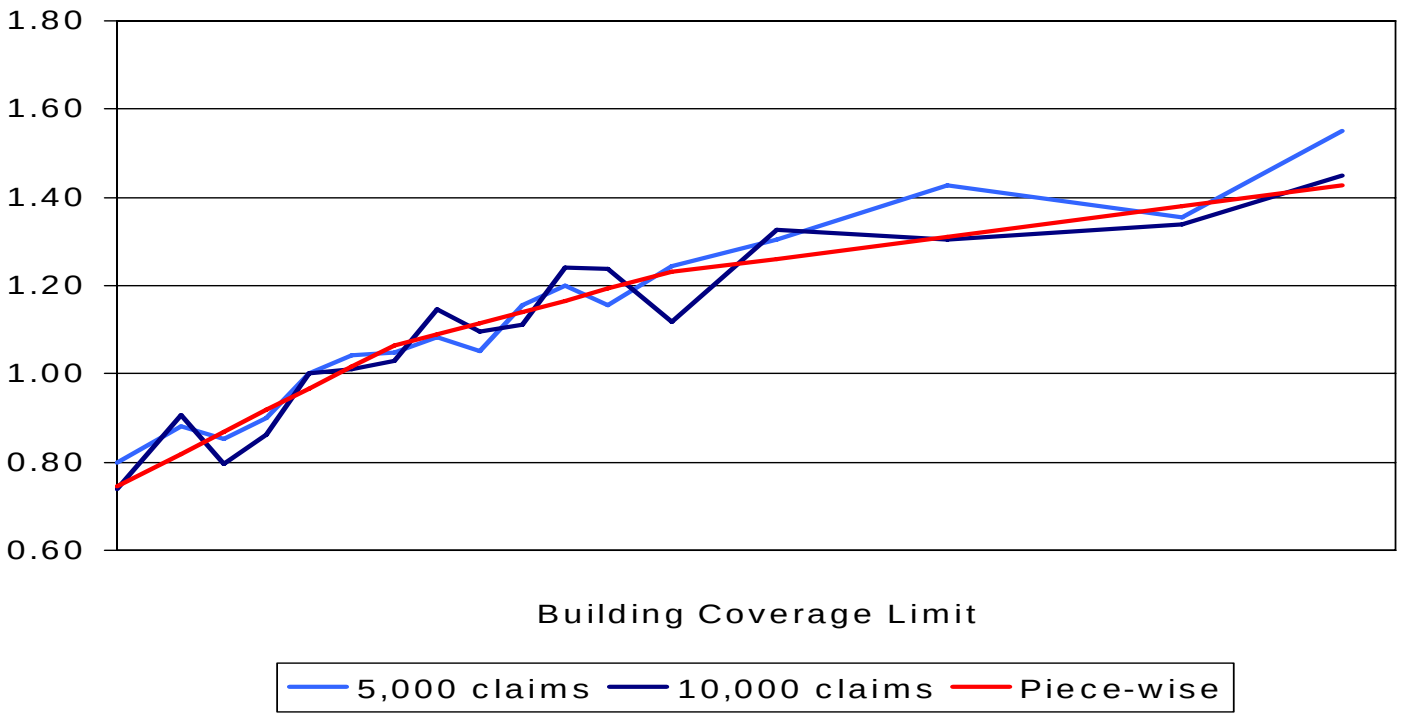
- You might consider splitting losses into multiple parts
  - By Peril/Cause of Loss
  - By size of loss
- Should give more homogeneous buckets to work with
- Even so, you still may end up with simple models



Watch the degrees of freedom

- Too many variables in a model, or too many levels
- Might need to treat some variables as continuous
- May want to look at piece-wise/spline approaches

ILLUSTRATIVE



Another alternative – analyze loss ratios

- Could consider if you don't trust claim counts, and don't have access to some rating variables
- One benefit — it's easy to combine coverages
- May be useful if your rating plan factors have been relatively stable over time
  - You may be able to adjust premiums of only a few variables that have been changed in the period
  - Otherwise need premiums at current rate level, which can be difficult to determine
- Will still have to deal with volatility in claim size
- Reasonable candidate for tier or underwriting model (e.g., for commercial, give guidance for schedule debits and credits)

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)
  - Make use of competitor filings to check magnitude and direction of results
- General understanding of modeling will help maximize value of your data
  - Data volume
  - Model design

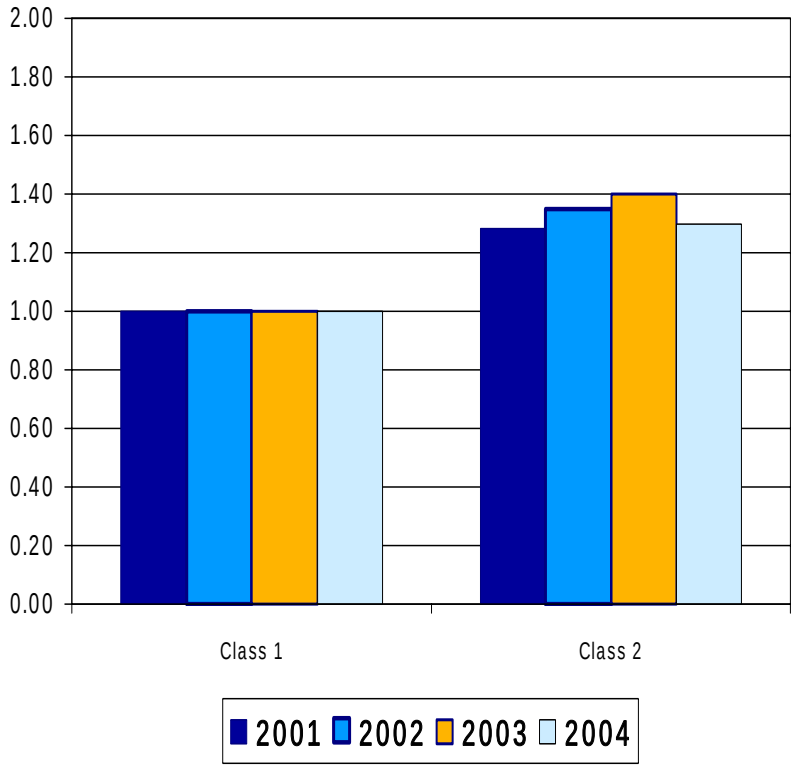
Dealing with volatility

- 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 or mandated discounts
  - Then let model compensate when estimating remaining variables

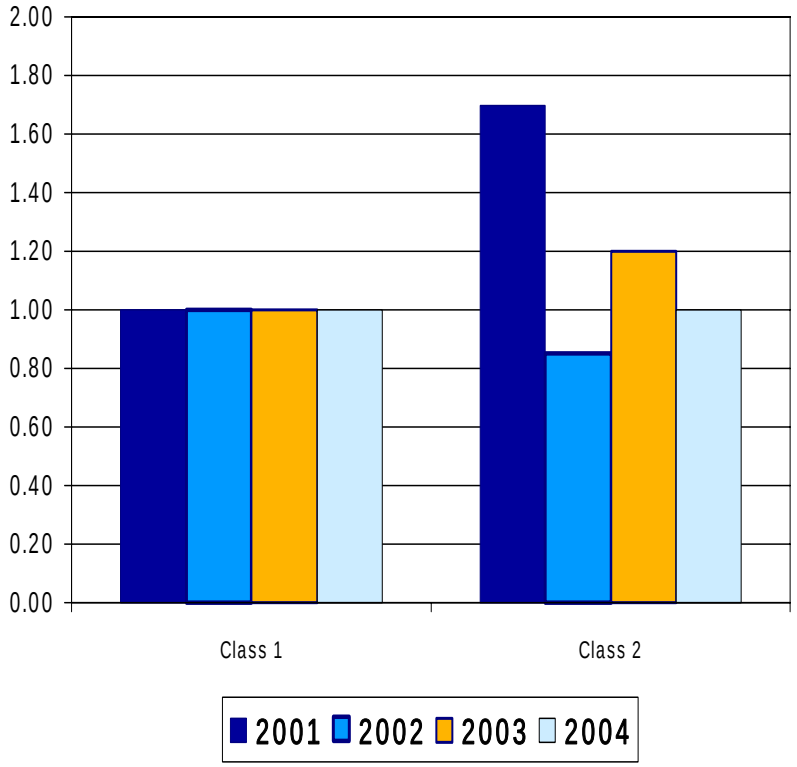
Consistency by year

ILLUSTRATIVE

Desirable



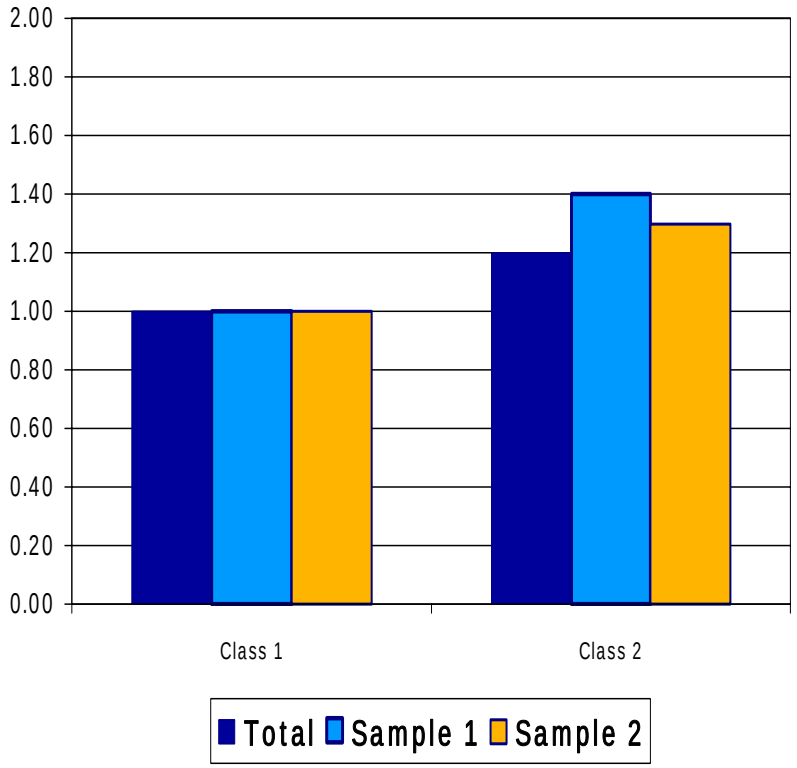
Not So Desirable



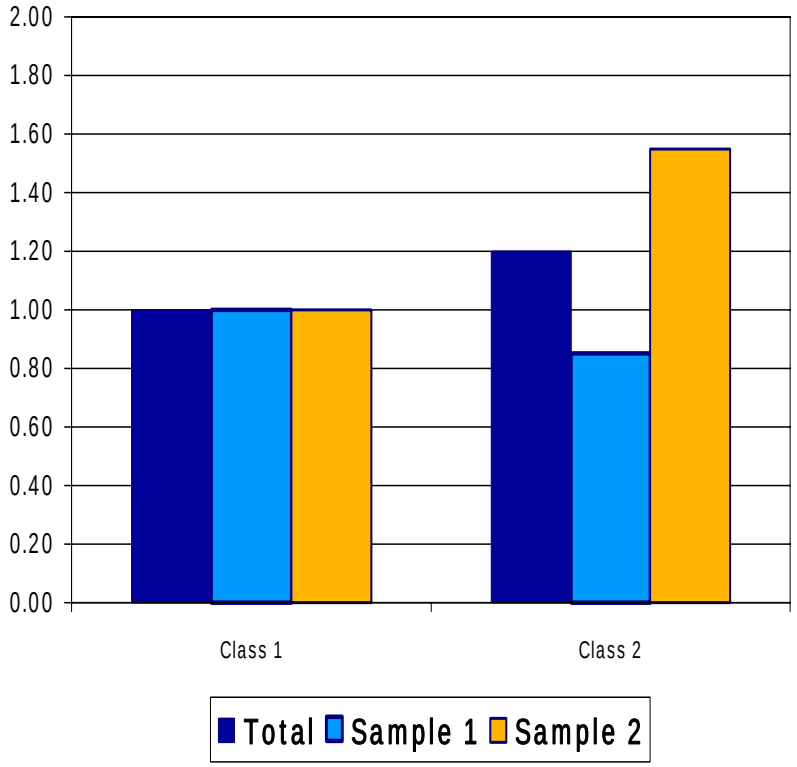
Consistency in samples

ILLUSTRATIVE

Desirable



Not So Desirable



Summary: modeling for small companies v. larger companies

- Keep in mind that modeling may be harder for smaller companies
- Don't try to do more than is realistic for the amount of data you have

Business Goal	Likelihood of Success
Develop full class plan	✓
Refine granularity for a given variable	✓✓✓✓
Explore new variables	✓✓✓
Explore variable interactions	✓✓—✓✓✓
Develop tier definitions/factors	✓✓✓
Develop underwriting models	✓✓✓
Analyze retention	✓✓✓✓

✓ Low likelihood ✓✓✓✓✓ High Likelihood (Large Company)

Closing thoughts

- Predictive modeling can help you make better business decisions
  - Even simple analyses can be better than traditional approaches
  - At a minimum, it can help you convince other stakeholders when contemplating something new
- Leverage what you've learned from PM
  - Develop monitoring / early warning reports
    - Monthly retention reports for desired/undesirable segments
    - Policies written or quotes made by segment
- Develop a strategy for the future
  - Plan to move beyond 'one-off' database construction
  - Identify data variables which you might have available while competitors do not
  - Make sure any newly identified variables are available for easy merge with traditional data