Claim Reserving: Performance Testing and the Control Cycle

Presentation at the CAMAR Spring Meeting

YI JING, FCAS, MAAA

June 9, 2010

TOWERS WATSON

Today's agenda

- Defining the problem
- The actuarial control cycle — embedding reserve risk management
- Performance testing — in general and in the context of reserves
- Case studies — real-world results

Questions for the reserving actuary

- **How do you know that the methods you are currently using are the "best"?**

  — What evidence supports your selection of methods?

  — What are the right weights for combining the results of the methods?

  — How do you decide when to change methods?

  — What is the confidence range around estimates?

  — How do you evaluate the cost/benefit of developing new data sources or implementing more complex methods?

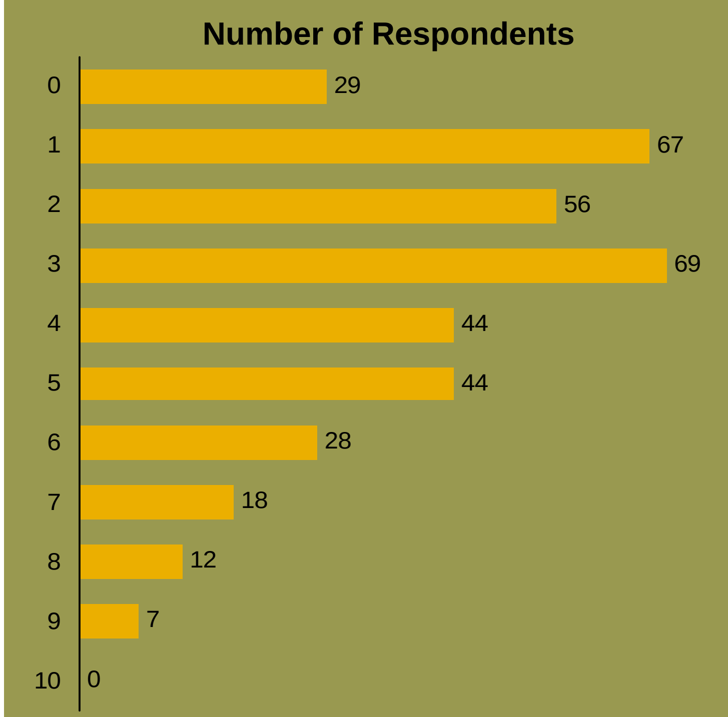
The results of our research illustrate the prevalence of actuarial overconfidence

Towers Watson Confidence Quiz

The Quiz

- Objective: To test respondents understanding of the **limits** of their knowledge
- Respondents were asked to answer ten questions related to their general knowledge of the global property/casualty industry
- For each answer, respondents were asked to provide a range that offered a 90% confidence interval that they would answer correctly
- Ideally (i.e., if “well calibrated”), respondents should have gotten nine out of ten questions correct

Raw Scores of Respondents



Note: based on 374 respondents as of 4/5/04.
Profile of respondents: 86% work in P/C industry; 73% are actuaries.

Reserves are forecasts!

- An actuarial method is used to produce a *forecast* of future claim payments
- An actuarial method consists of
  - An algorithm
  - A data set
  - A set of intervention points

$$\hat{L}_m^{(t)} = m(a, d, p)$$
$$L^{(t)} = m(a, d, p) + \varepsilon_m$$

- The actuary must

1. Choose a finite set of methods from the universe *M*

$$\{m_1, m_2, \ldots m_n\}$$

2. Choose a set of weights to combine the results of each method together

$$\{w_1, w_2, \ldots w_n\}$$

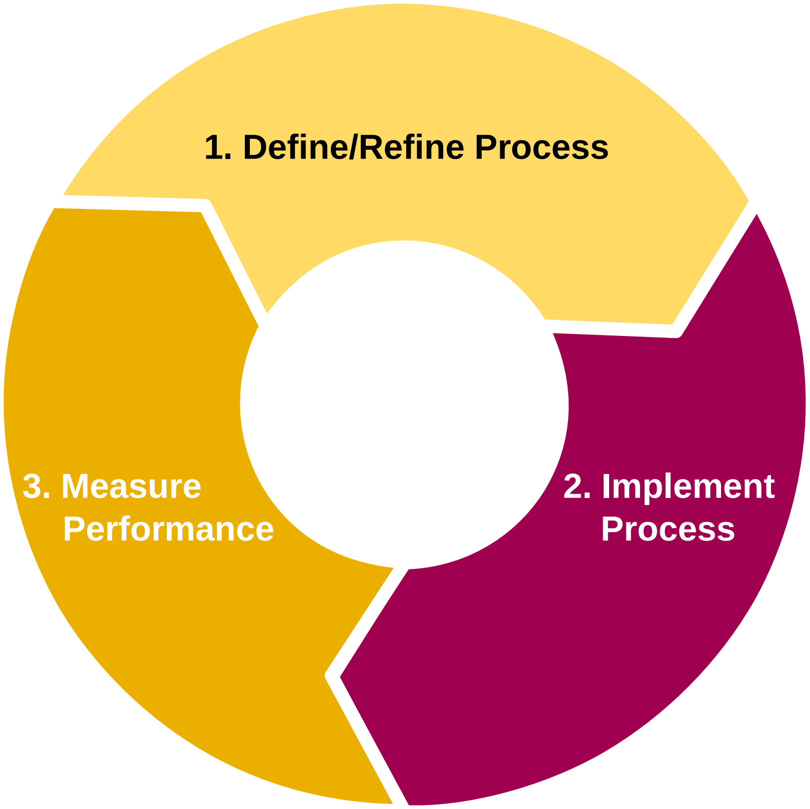
- *Performance testing, via a formal control cycle, can help the actuary make these choices in a rigorous manner*

Performance testing of reserving methods can be part of an institutionalized control cycle

The Actuarial Control Cycle for the Reserving Process
Embedding Reserve Risk Management

Formal Performance Testing

- Are the current methods appropriate? Would changes to methods improve estimation skill?
- Are the data and other input accurate and sufficient? Would improvements or expansion of data improve estimation skill?
- Are there opportunities to improve process flow?
- Are emerging estimation errors within tolerances?

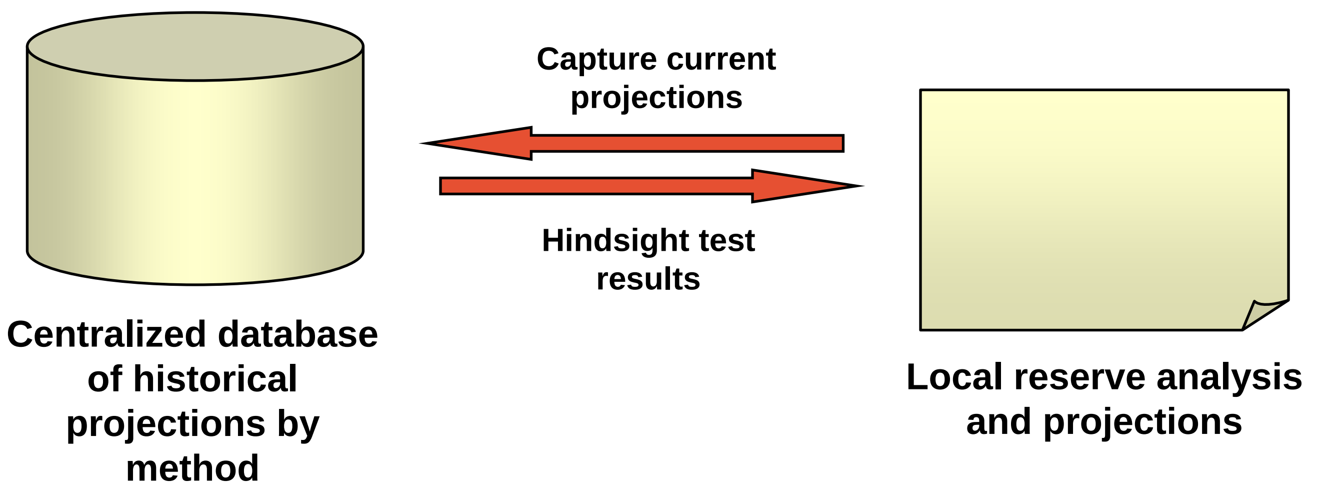


Reserving Process Elements

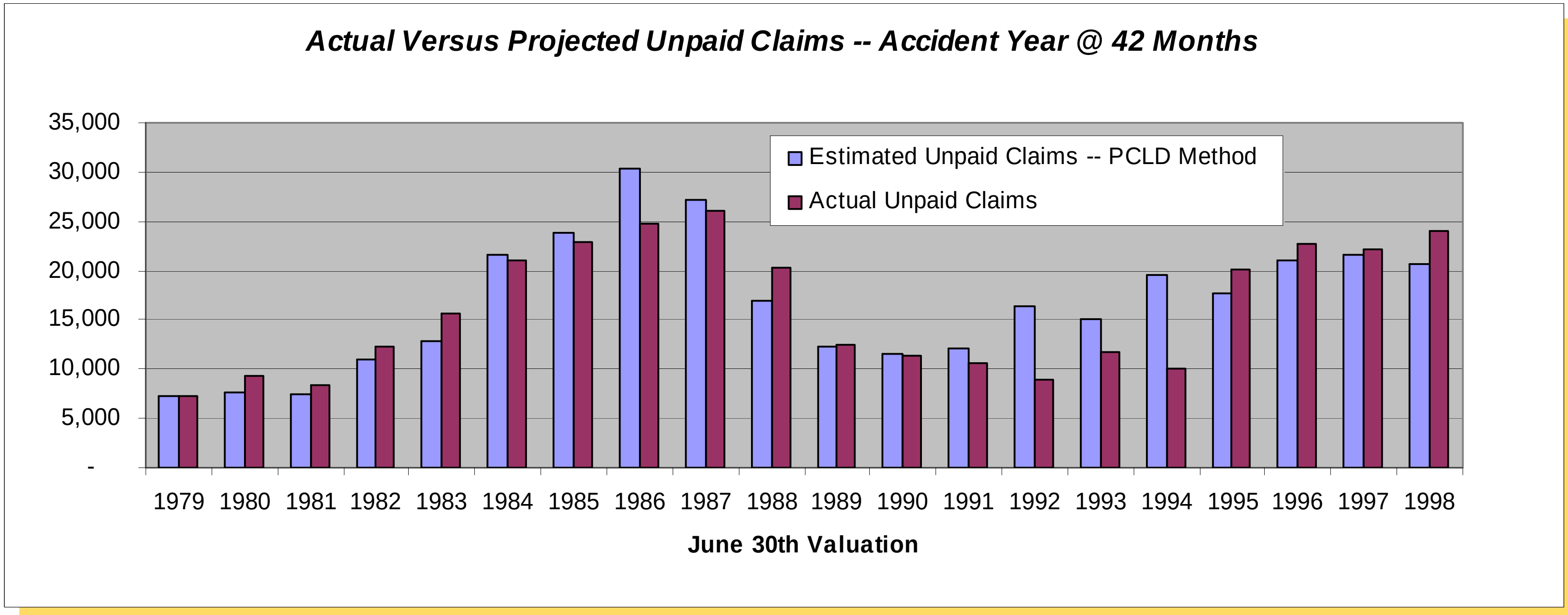
- Data used
- Actuarial methods employed
- Operational input
- Judgments and intervention points
- Process flow and timeline
- Quality assurance process

We have already worked with one major client to install performance testing and a control cycle

- Corporate Actuary responsible for reserves set by decentralized organization of actuaries within each business unit
- Standard templates and database used to capture quarterly projections on an ongoing basis
- Actuaries review performance test results prior to each quarterly reserve-setting exercise; perform more detailed analysis annually



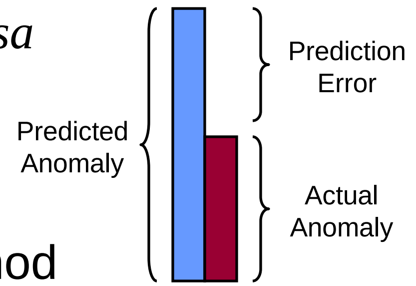
Performance testing is a formal analysis of prediction errors

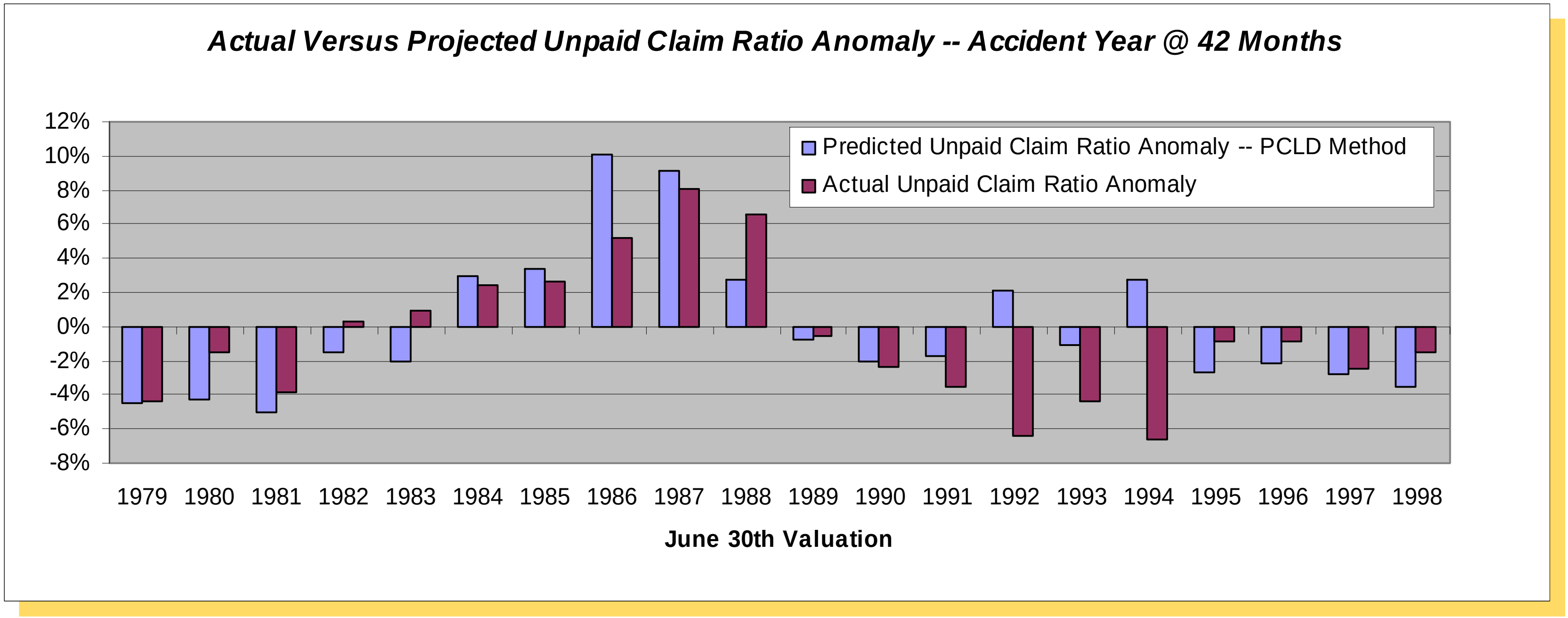


- Test a particular method by looking at historical performance – comparing estimates from the method with actual run-off
- Performance testing is a formalized process, not just a numerical exercise

Performance testing yields a formal measure of skill

- The skill of a method is measured by: $Skill_m = 1 - mse_m / msa$
  - mse = mean squared error
  - msa = mean squared anomaly
- Skill is the proportion of variance “explained” by the method

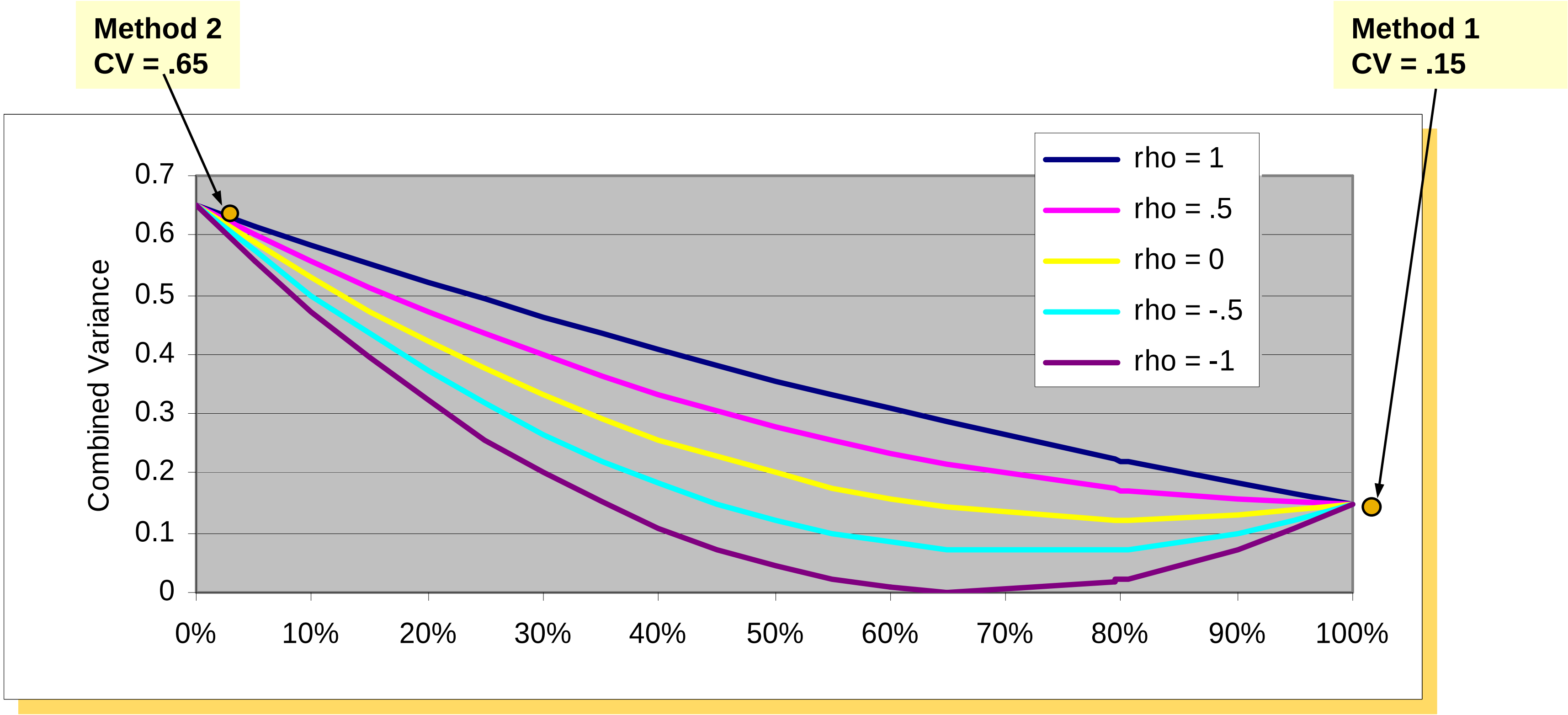




Calculation of PCLD skill at 42 months

Valuation Date (1)	Accident Year (2)	Earned Premium (3)	PCLD Projected Unpaid (4)	Ratio (5)	"Actual" Unpaid (6)	Ratio (7)	Percent Paid (8)	Actual Anomaly (9)	Predicted Anomaly (10)	Error (11)
Jun-79	1976	59,040	7,173	12.1%	7,211	12.2%	97%	-3.73%	-3.79%	-0.06%
Jun-80	1977	61,516	7,598	12.4%	9,268	15.1%	100%	-0.88%	-3.59%	-2.71%
Jun-81	1978	65,088	7,506	11.5%	8,295	12.7%	100%	-3.20%	-4.41%	-1.21%
Jun-82	1979	72,444	10,910	15.1%	12,214	16.9%	99%	0.92%	-0.88%	-1.80%
Jun-83	1980	88,861	12,917	14.5%	15,613	17.6%	100%	1.63%	-1.41%	-3.03%
Jun-84	1981	110,447	21,609	19.6%	20,989	19.0%	100%	3.06%	3.62%	0.56%
Jun-85	1982	118,920	23,766	20.0%	22,899	19.3%	100%	3.31%	4.04%	0.73%
Jun-86	1983	113,763	30,334	26.7%	24,796	21.8%	100%	5.85%	10.72%	4.87%
Jun-87	1984	105,594	27,098	25.7%	26,040	24.7%	100%	8.72%	9.72%	1.00%
Jun-88	1985	87,695	16,932	19.3%	20,342	23.2%	100%	7.25%	3.37%	-3.89%
Jun-89	1986	77,969	12,301	15.8%	12,504	16.0%	99%	0.09%	-0.17%	-0.26%
Jun-90	1987	79,207	11,550	14.6%	11,289	14.3%	100%	-1.69%	-1.36%	0.33%
Jun-91	1988	81,725	12,157	14.9%	10,683	13.1%	100%	-2.87%	-1.07%	1.80%
Jun-92	1989	87,874	16,463	18.7%	8,905	10.1%	100%	-5.81%	2.79%	8.60%
Jun-93	1990	96,657	15,031	15.6%	11,794	12.2%	100%	-3.74%	-0.39%	3.35%
Jun-94	1991	101,162	19,541	19.3%	10,108	10.0%	100%	-5.95%	3.37%	9.32%
Jun-95	1992	128,231	17,767	13.9%	20,104	15.7%	97%	-0.27%	-2.09%	-1.82%
Jun-96	1993	145,262	20,967	14.4%	22,764	15.7%	94%	-0.27%	-1.51%	-1.24%
Jun-97	1994	156,751	21,633	13.8%	22,086	14.1%	93%	-1.85%	-2.14%	-0.29%
Jun-98	1995	159,211	20,733	13.0%	23,924	15.0%	88%	-0.92%	-2.92%	-2.00%
			Weighted Average Ratios	16.6%		15.9%				
			Bias	0.6%						
								msa =		0.1573%
								mse =		0.1249%
								Skill =		21%

The minimum-variance weighting of methods depends on their variances and their correlation



- For a given correlation, the optimal weights are those with the minimum combined variance
  - Minimum starts at the very right, when correlation is 100%
  - Minimum gradually shifts leftward as correlation decreases

Optimal weights for any set of methods can be found that minimize the variance of the weighted average estimate

- Assume two methods X and Y are weighted with weights a and b, $U = a \times X + b \times Y$
- Variance of the weighted estimate U:

$$V_U = a^2 \times V_X + 2 \times a \times b \times Cov(X,Y) + b^2 \times V_Y$$
$$= a^2 \times V_X + 2 \times a \times b \times \rho_{X,Y} \times \sigma_X \times \sigma_Y + b^2 \times V_Y$$

- To minimize V_U, taking the derivative of V_U with respect to a, setting it equal to 0, and solving for a, we get:

$$a = \frac{V_Y - \rho_{X,Y} \times \sigma_X \times \sigma_Y}{V_X - 2 \times \rho_{X,Y} \times \sigma_X \times \sigma_Y + V_Y}$$

$$b = 1 - a = \frac{V_X - \rho_{X,Y} \times \sigma_X \times \sigma_Y}{V_X - 2 \times \rho_{X,Y} \times \sigma_X \times \sigma_Y + V_Y}$$

- If X and Y are uncorrelated, $\rho_{X,Y} = 0$, and if X and Y have the same variance, then $a = b = 50\%$, if $V_X < V_Y$, $a > b$, more weight is given to method X

- Math can easily be extended to more than two methods, implemented using Solver

Case Study: U.S. Insurer

- Commercial Auto BI data
  - 1972 to 1998 accident years – June 30th valuations
  - Paid and incurred counts and amounts
  - Estimates of claim liabilities from 1979 to 1998 – twenty years
  - December 31st valuation used as "actual ultimate"
- Environmental influences during the period add difficulty to estimation
  - Economic and social inflation
  - Operational changes in claim department
  - Changes in underwriting posture

Formally testing alternative methods yields some interesting and counterintuitive results

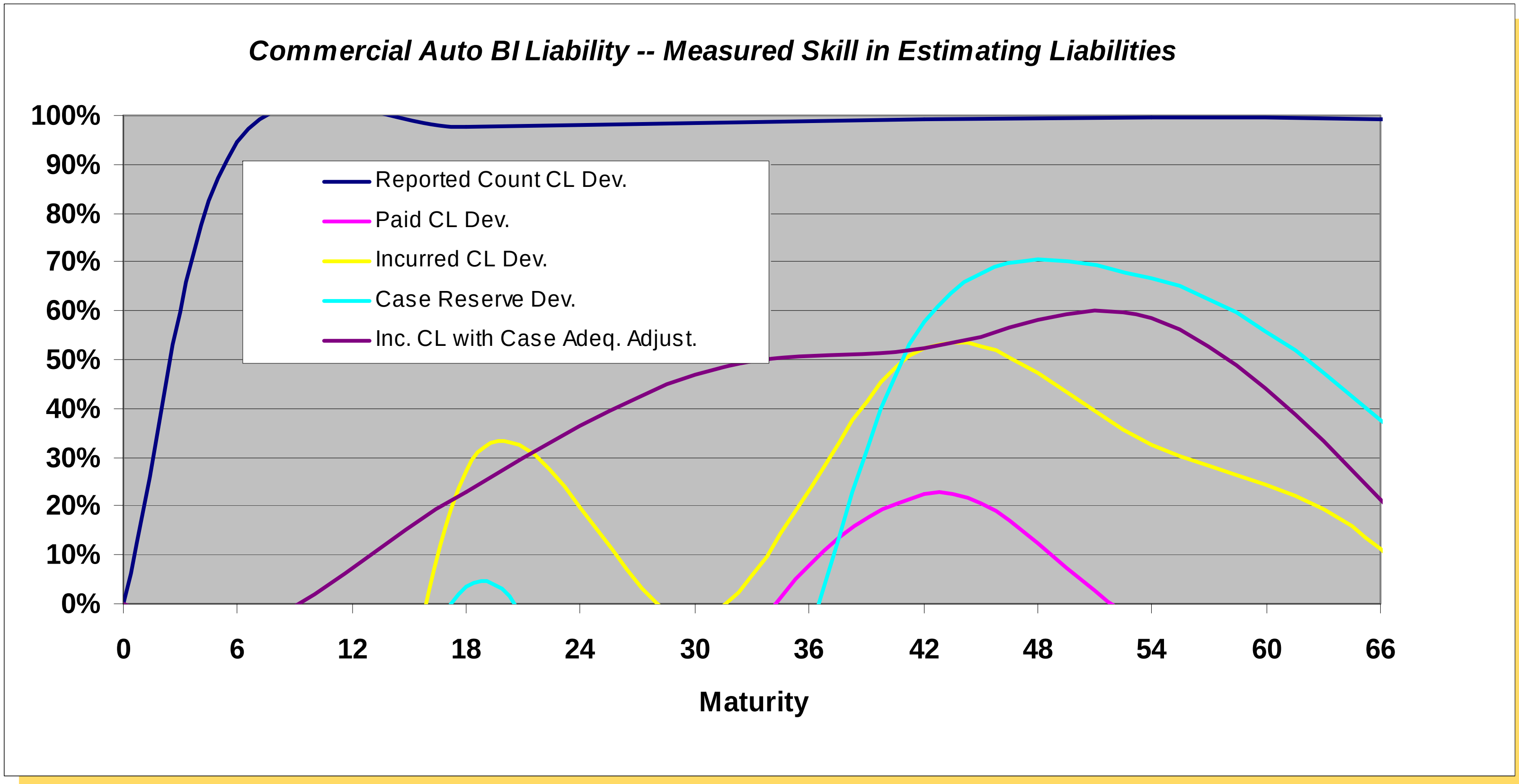
- Sometimes projecting case reserves is the best method
- Methods that use claim counts and averages may outperform
- Methods that formally adjust for changing claim settlement rates or changing case reserve adequacy can produce better estimates
- The degree of correlation between methods is an important consideration in selecting methods, and weights used to combine them

Actuarial methods subjected to performance testing

Actuarial Projection Method	Skill for Accident Year @ 42 Months	Overall Skill – for Latest Ten Accident Years
Paid Chain-Ladder	21%	13%
Reported Chain-Ladder	52%	32%
Case Reserve Development	60%	22%
Reported Count Chain-Ladder	99%	99%
Case Adequacy Adjusted Reported Chain-Ladder	52%	52%

- Note that absolute level of skill results are low due to changing case reserve adequacy and claim settlement patterns

Skill varies by maturity, suggesting that methods should vary by maturity



- Note that skill can be negative (e.g., paid loss projection method at 6 months), implying that it induces volatility rather than explaining it

The correlation among the errors is relatively high, suggesting an underlying systematic error affecting all methods

- Combining methods together to get a best estimate is most effective when the estimates are independent, so correlation is not a good thing
- Count projections are less correlated, suggesting that a counts-and-averages method might be worth introducing
- Reported loss and case reserve projections are particularly correlated

Correlation of Liability Estimation Errors

	RC	PL	RL	OL	ARL	AOL	ELR
Reported Counts (RC)	100%	61%	36%	38%	57%	48%	19%
Paid Loss (PL)		100%	72%	75%	95%	87%	72%
Reported Loss (RL)			100%	98%	77%	85%	69%
Case Outstanding Loss (OL)				100%	81%	92%	74%
Adj Reported Loss (ARL)					100%	94%	79%
Adj Case O/S Loss (AOL)						100%	82%
3-Year ELR (ELR)							100%

Good reasons to do performance testing

1. Opportunity to improve accuracy of estimates
2. Formal rationale for selected actuarial methods
3. Input to development of reserve ranges
4. Manage actuarial overconfidence
5. Cost / benefit of enhancements to data and systems
6. Supports Solvency II / Economic Capital
   - Embeds reserve risk management
   - Empirical validation of stochastic reserve risk models