THE MACK/MURPHY MODEL FRAMEWORK: FROM THEORY TO PRACTICE

CLFM VARIANCE ESTIMATES

CLRS 2011

Daniel Murphy

Trinostics LLC

TRINOSTICS LLC

Variance formulas for estimates of the first future diagonal are relatively painless

$$Y = f \cdot X + X^{\alpha/2}\sigma e \quad (1)$$
$$= expectation\,term + error\,term$$

- **Parameter risk** is a characteristic of the *expectation term* and arises from the error surrounding the estimation of the parameter f
- **Process risk** is a characteristic of the *error term* and arises from the random nature of the processes that cause claims development outcomes to differ from their mean value
- The formula for $Var(\hat{f})$ in the regression model (1) is well known*
  – So the variance of the *expectation term* is $X^2 Var(\hat{f})$ because X is a known value
- An estimate of the value of σ^2 in the regression model (1) is well known*
  – So the variance of the *error term* is $X^{\alpha}\widehat{\sigma^2}$ because X is a known value

* see Excel method on slide 12

Variance formulas for estimates beyond the first future diagonal are calculated recursively

$$\hat{Y} = \hat{f} \cdot \hat{X} + \hat{X}^{\alpha/2} \sigma e$$

- The second future diagonal is the product of the first future diagonal (an estimate) and a development factor (an estimate)
- **Parameter risk** comes from the formula for the variance of the product of two estimates
  - Straightforward, but complicated
- The formula for **Process risk** is derived via conditional expected values and can be difficult when α is not a positive integer
  - We approximate with linear interpolation for α between two integers
  - For negative values of α we approximate with random sampling
- We will not explore those details today
- For the remainder of this presentation we will focus on applying these techniques to real data (Dan) and simulated data (Tim)

Schedule P Date on CAS Website

- Glenn Meyers and Peng Shi combined 12/31/1997 Schedule P data with data from one or more subsequent Schedule P's, and posted six lines of business on the CAS website http://www.casact.org/research/index.cfm?fa=loss_reserves_data
- Group data is presented at the group level (202 observations for Commercial Auto)
- Predictions of expected outcomes (point estimates) and variability (stochastic estimates) based on the 12/31/1997 historical triangles can be compared with the actual development experience through 10 years of maturity
- We investigated paid loss chain ladder projections for Commercial Auto and compared them with the actual future values
- We investigated the variability of those projections by looking at one-standard-error (SE=√MSE) upper bounds (one-sided confidence intervals) based on the Mack/Murphy and CLFM stochastic methods

Commercial Auto Paid Loss Schedule P Data
202 companies/groups, aggregated, unedited

AY	Age (years) 1	2	3	4	5	6	7	8	9	10
1988	403,897	879,615	1,320,001	1,575,326	1,726,769	1,806,048	1,844,129	1,865,130	1,874,390	1,881,860
1989	486,416	1,105,008	1,569,042	1,868,765	2,030,776	2,117,440	2,159,463	2,181,925	2,194,988	2,181,183
1990	507,484	1,129,279	1,633,693	1,935,812	2,100,597	2,189,875	2,234,113	2,250,967	2,240,741	2,241,430
1991	477,344	1,050,305	1,482,563	1,757,305	1,916,740	1,997,076	2,029,720	2,036,833	2,046,698	2,052,861
1992	499,400	1,096,270	1,545,385	1,830,931	2,002,076	2,080,371	2,093,979	2,112,494	2,121,518	2,133,179
1993	547,793	1,190,155	1,566,794	1,989,390	2,145,398	2,198,677	2,237,416	2,256,725	2,269,487	2,275,536
1994	643,215	1,342,679	1,882,014	2,213,787	2,376,925	2,472,365	2,516,469	2,532,667	2,546,490	2,556,633
1995	671,282	1,419,751	1,884,353	2,181,464	2,350,785	2,451,149	2,490,801	2,513,996	2,532,394	2,542,675
1996	761,287	1,406,276	1,865,804	2,185,616	2,384,848	2,486,224	2,546,224	2,571,755	2,580,461	2,587,707
1997	665,573	1,330,943	1,851,447	2,244,620	2,454,753	2,571,428	2,639,771	2,668,276	2,681,447	2,681,165

- Pretend you only know the triangle of black numbers
- You can predict
  - the values in the age 10 column (tan and gray), or
  - the values on the next diagonal (the tan numbers), or …
- You wish to estimate the variability of outcomes around your prediction
- For various reasons we will look at just a subset of 63 groups/companies

Trinostics LLC

For 63 selected companies/groups, we compare predicted vs. actual age 2 values for AY 1997

Industry Aggregate
63 observations

AY	Age (years) 1	2	Age-to-Age Factors
1988	88,235	187,772	2.128
1989	157,893	352,458	2.232
1990	160,275	371,405	2.317
1991	147,802	351,159	2.376
1992	146,135	332,687	2.277
1993	145,642	312,577	2.146
1994	191,203	382,485	2.000
1995	206,719	477,660	2.311
1996	288,253	454,348	1.576
1997	205,258	*401,159*	*1.954*

unknown as of 12/97

	Averages
Weighted	2.103
Simple	2.152
4-Year	1.956

- The 63 groups make up about 31% of the industry
- Our goal will be to predict the age 2 value for AY 1997
- If we had selected the weighted average of accident years 1993-1996 we would have been almost perfect
- We will select a development factor of 1.954 *for every company's chain ladder method*
- That model will be unbiased because the mean of the predictions will equal the true mean value

Four companies/groups selected at random

Group A

AY	Age (years) 1	2	RTR factor	
1988	34	54	1.588	
1989	14	47	3.357	
1990	25	53	2.120	
1991	37	59	1.595	
1992	11	28	2.545	
1993	10	18	1.800	
1994	15	30	2.000	
1995	15	25	1.667	
1996	26	37	1.423	
1997	25	56		
		49	1.954	4.523

Group B

AY	Age (years) 1	2	RTR factor	
1988	450	414	0.920	
1989	193	454	2.352	
1990	310	529	1.706	
1991	186	428	2.301	
1992	131	370	2.824	
1993	96	244	2.542	
1994	105	215	2.048	
1995	101	177	1.752	
1996	123	324	2.634	
1997	89	133		
		174	1.954	1.312

Actual outcome

Expected outcome

Group C

AY	Age (years) 1	2	RTR factor	
1988	1,938	4,886	2.521	
1989	2,617	6,094	2.329	
1990	2,371	5,322	2.245	
1991	1,524	3,465	2.274	
1992	792	1,821	2.299	
1993	616	1,282	2.081	
1994	708	1,247	1.761	
1995	867	1,120	1.292	
1996	953	1,305	1.369	
1997	1,036	1,929		
		2,024	1.954	2.622

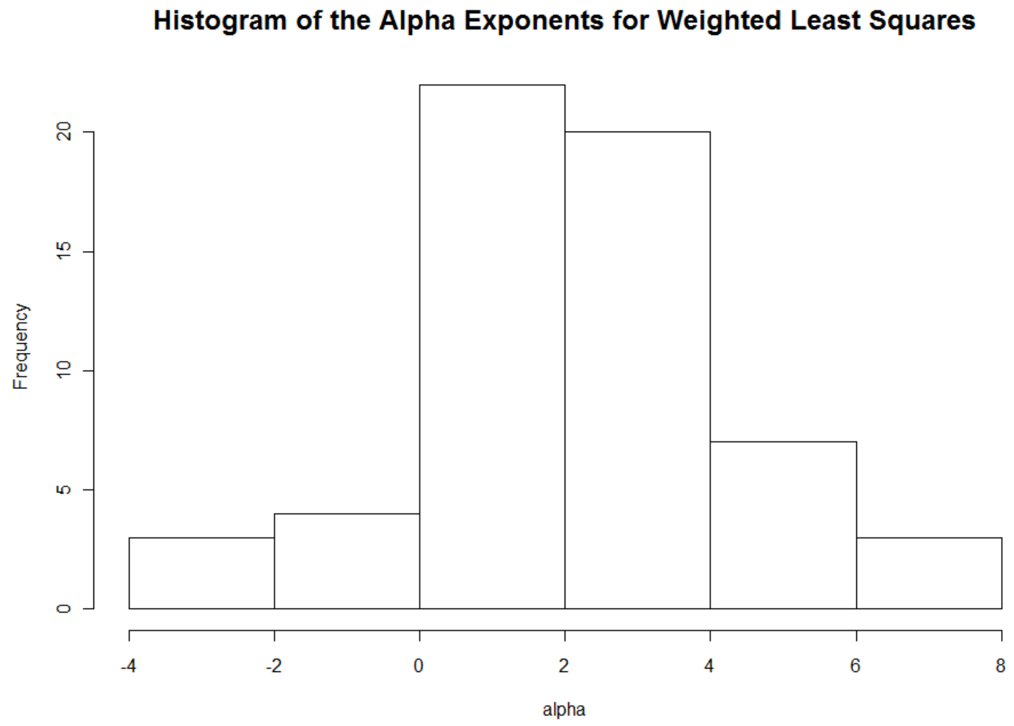
Group D

AY	Age (years) 1	2	RTR factor	
1988	13,457	34,825	2.588	
1989	19,839	35,763	1.803	
1990	12,539	39,729	3.168	
1991	10,985	43,846	3.991	
1992	8,211	47,369	5.769	
1993	13,549	28,099	2.074	
1994	15,013	31,542	2.101	
1995	18,861	36,113	1.915	
1996	25,002	44,377	1.775	
1997	27,309	61,134		
		53,362	1.954	-2.147

alpha

- The predicted value in red is based on the selected 1.954 link ratio
- The alpha value in blue is the exponent of the Age 1 losses in the weighted least squares, zero-intercept regression model (CLFM) whose best estimate of the slope factor is 1.954

Distribution of alpha's across 63 applications of CLFM using same selected 1.954 LDF



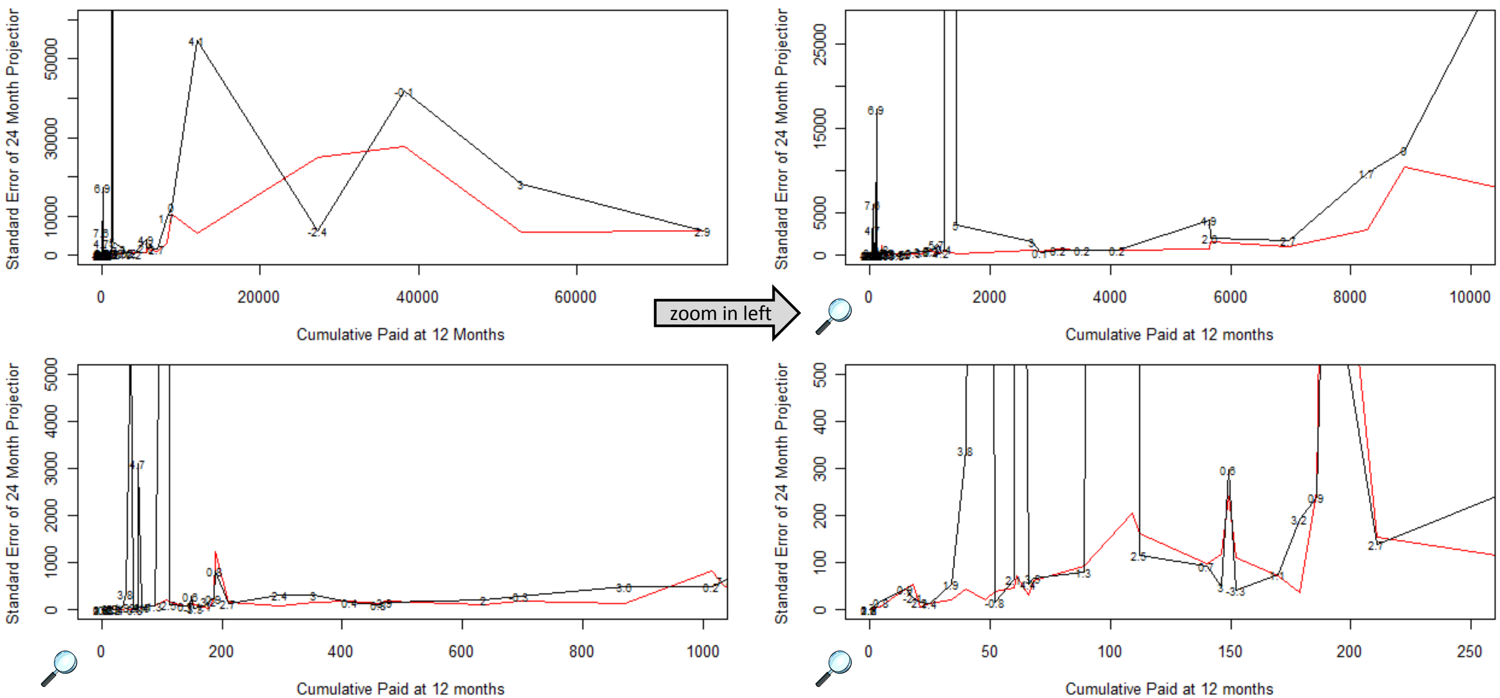
- Most of the alphas are concentrated around the weighted average (α=1) and simple average (α=2) exponents

Expected age 2 values vs. actual age 2 values
Relative to 1 SE upper bounds on outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	GRP	Age 1 Value	Actual Age 2 Value	CLFM Expected @ age 2 1.954*(2)	CLFM Parameter Risk	CLFM Process Risk	√(5)²+(6)² CLFM Total Risk (SE_{CLFM})	One-Sided Conf Int (4) + (7)	(3) > (8)?	MM (wtd) Total Risk (SE_{MM})	One- Sided Conf Int (4) + (10)	(3) > (11)?	(7) > (10)?	MM Expected @ age 2	MM bias (14) - (4)	MM SE incl bias √(10)²+(15)²	One- Sided Conf Int (4) + (16)	(3) > (17)?
1		4	4	8	1	14	14	21	FALSE	6	13	FALSE	TRUE	8	0	6	13	FALSE
2		18	27	35	25	8	27	62	FALSE	54	89	FALSE	FALSE	55	20	58	93	FALSE
3		21	53	41	5	15	16	57	FALSE	6	47	TRUE	TRUE	35	-6	8	49	TRUE
4	A	66	103	129	11	55	56	185	FALSE	31	160	FALSE	TRUE	91	-38	49	178	FALSE
5		68	98	133	19	64	67	200	FALSE	62	194	FALSE	TRUE	151	19	64	197	FALSE
6	B	89	133	174	20	76	79	253	FALSE	93	267	FALSE	FALSE	166	-8	93	267	FALSE
7		152	347	297	23	37	43	340	TRUE	110	407	FALSE	FALSE	396	99	148	445	FALSE
8		170	423	332	32	71	78	410	TRUE	72	404	TRUE	TRUE	327	-5	72	404	TRUE
9		179	264	350	31	199	202	552	FALSE	37	387	FALSE	TRUE	260	-90	97	447	FALSE
10		461	926	901	71	117	137	1,038	FALSE	142	1,043	FALSE	FALSE	904	3	142	1,043	FALSE
11		471	1,206	921	77	144	163	1,084	TRUE	181	1,102	TRUE	FALSE	809	-111	213	1,133	TRUE
12		636	2,202	1,243	96	207	228	1,471	TRUE	113	1,356	TRUE	TRUE	1,199	-44	121	1,364	TRUE
13	C	1,036	1,929	2,025	152	447	472	2,497	FALSE	494	2,519	FALSE	FALSE	2,220	195	531	2,556	FALSE
14		1,293	2,101	2,527	993	9,143,615	9,143,615	9,146,142	FALSE	499	3,026	FALSE	TRUE	2,261	-266	565	3,092	FALSE
15		4,132	6,457	8,076	272	606	664	8,740	FALSE	697	8,773	FALSE	FALSE	8,097	21	697	8,773	FALSE
16		5,003	10,987	9,778	472	3,001	3,038	12,815	FALSE	1,108	10,886	TRUE	TRUE	10,488	710	1,316	11,094	FALSE
17		5,645	8,694	11,033	269	5,776	5,782	16,815	FALSE	720	11,753	FALSE	TRUE	9,788	-1,244	1,438	12,470	FALSE
18	D	27,309	61,134	53,373	4,456	5,339	6,955	60,328	TRUE	24,936	78,309	FALSE	FALSE	67,880	14,507	28,849	82,222	FALSE
19		52,893	98,038	103,375	4,390	17,101	17,656	121,031	FALSE	5,893	109,268	FALSE	TRUE	113,905	10,530	12,067	115,441	FALSE
Tot	63	205,258	401,159	401,159					5			5	37	423,751	22,592			4
Avg			6,367.6	6,367.6					8%			8%	59%	6,726.2				6%

- The average prediction equals the average industry value, ∴ method is unbiased
- CLFM standard error > MM standard error 60% of the time
- MM+bias method appears to predict a wider spread
- MM+bias method has not been extensively studied to presenters' knowledge

Graph of 63 Standard Error Estimates Mack/Murphy vs. CLFM



- CLFM SE is in black, Mack/Murphy SE is in red
- Most of the time CLFM SE > M/M SE, but not always
- The CLFM SE can be volatile: depends on the volatility of the underlying triangle and the proximity of the selected factor to historical experience

Summary of CLFM variance analysis of Commercial Auto Schedule P data on CAS website

- The CLFM standard errors were greater than the M/M standard errors 60% of the time
- The standard errors do not appear understated, but hard to tell
  - 6-8% of the groups' outcomes exceeded one standard error above the mean
  - 16% would be indicated with a bell shaped distribution
- More data points (!) and more study is needed
  - Only 63 observations studied here
  - Entire industry of 202 groups in the data base is less than desirable
- Historical data can have certain limitations that simulated data can sometimes overcome
  - Economic Scenario Generators
  - Claims process simulators (Take it over Tim!)

Calculating parameter, process risk with Excel

Group C alpha = 2.622

Transform data to form homoscedastic model

AY	X	Y	x= X/(X^alpha/2)	y= Y/(X^alpha/2)
1988	1,938	4,886	0.0949	0.2393
1989	2,617	6,094	0.0864	0.2013
1990	2,371	5,322	0.0891	0.2001
1991	1,524	3,465	0.1023	0.2325
1992	792	1,821	0.1254	0.2883
1993	616	1,282	0.1356	0.2822
1994	708	1,247	0.1298	0.2287
1995	867	1,120	0.1219	0.1575
1996	953	1,305	0.1184	0.1621
1997	1036			

LINEST OUTPUT

f =	1.954	0	
se(f) =	0.14684	#N/A	
	0.956792	0.049734	= σ
	177.1524	8	
	0.438178	0.019788	

Var(f) = se²(f) = 0.02156

σ^2 = 0.00247

parameter risk² = 23142.25 = 1036^2*Var(f)

parameter risk = 152

process risk² = 199591.9 =1036^alpha*σ²

process risk = 447

TRINOSTICS LLC

Daniel Murphy, FCAS, MAAA
dmurphy@trinostics.com

Email to receive final presentation slides