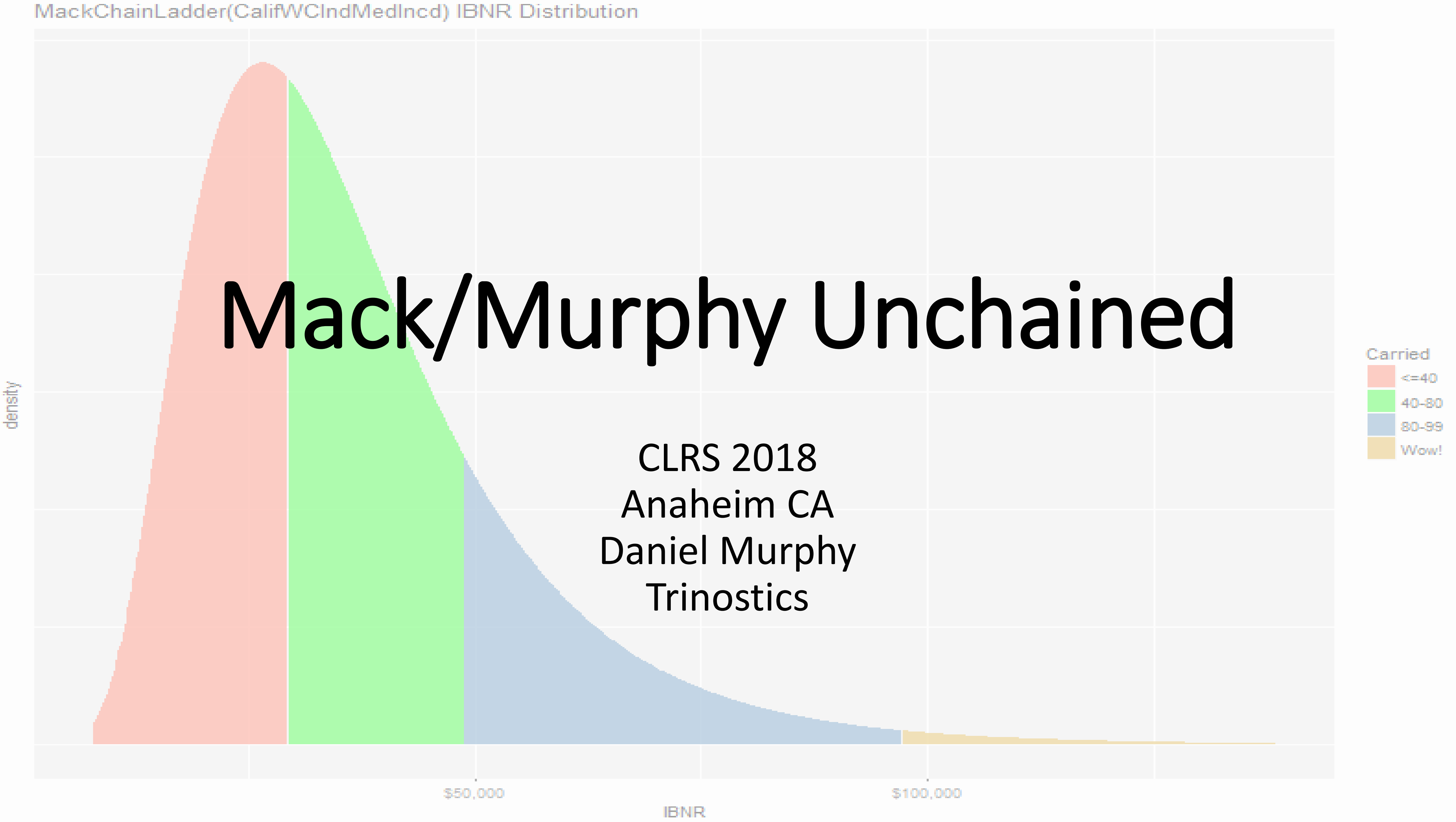
Mack/Murphy Unchained

CLRS 2018
Anaheim CA
Daniel Murphy
Trinostics



Strats

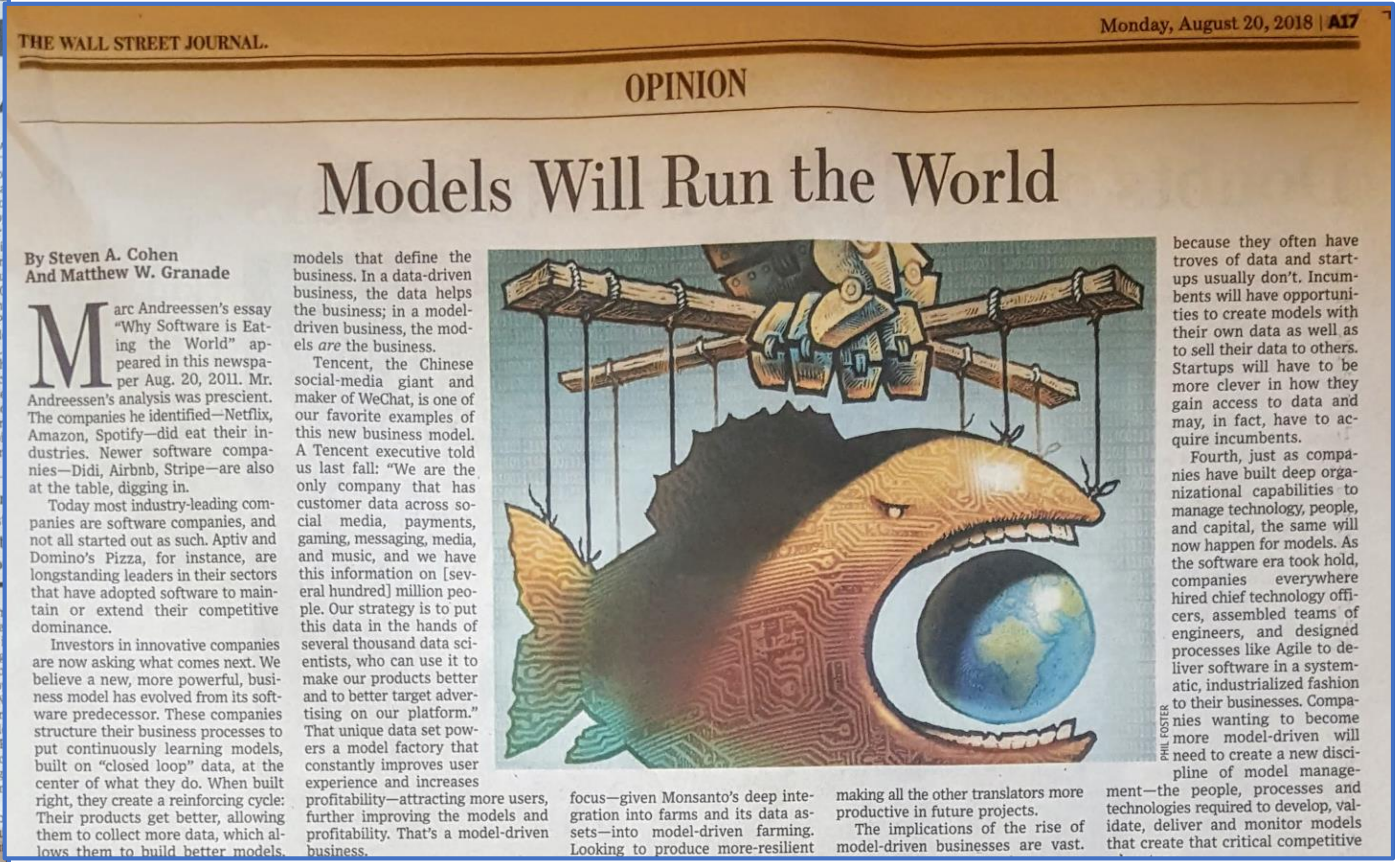
BUSINESS | FINANCE | TECHNOLOGY | MANAGEMENT THE WALL STREET JOURNAL. **** SATURDAY/SUNDAY

DJIA 25669.32 ▲ 110.59 0.4% NASDAQ 7816.33 ▲ 0.1% STOXX 600 381.06 ▼ 0.1% 10-YR. TREAS. ▼ 1/32, yield 2.873% OIL $65.91 ▲ $0.45 GOLD $1,176.50 ▲ $0.30

Trading Places

Wall Street used to have a strict hierarchy: Traders made money and won glory while programmers wrote code and stayed out of sight. Now, the line between 'the jocks' and 'the nerds' is disappearing.

In a model-driven business, the models *are* the business



THE WALL STREET JOURNAL. Monday, August 20, 2018 | A17

OPINION

Models Will Run the World

By Steven A. Cohen And Matthew W. Granade

Marc Andreessen's essay "Why Software is Eating the World" appeared in this newspaper Aug. 20, 2011. Mr. Andreessen's analysis was prescient. The companies he identified—Netflix, Amazon, Spotify—did eat their industries. Newer software companies—Didi, Airbnb, Stripe—are also at the table, digging in.

Today most industry-leading companies are software companies, and not all started out as such. Aptiv and Domino's Pizza, for instance, are longstanding leaders in their sectors that have adopted software to maintain or extend their competitive dominance.

Investors in innovative companies are now asking what comes next. We believe a new, more powerful, business model has evolved from its software predecessor. These companies structure their business processes to put continuously learning models, built on "closed loop" data, at the center of what they do. When built right, they create a reinforcing cycle: Their products get better, allowing them to collect more data, which allows them to build better models. models that define the business. In a data-driven business, the data helps the business; in a model-driven business, the models *are* the business.

Tencent, the Chinese social-media giant and maker of WeChat, is one of our favorite examples of this new business model. A Tencent executive told us last fall: "We are the only company that has customer data across social media, payments, gaming, messaging, media, and music, and we have this information on [several hundred] million people. Our strategy is to put this data in the hands of several thousand data scientists, who can use it to make our products better and to better target advertising on our platform." That unique data set powers a model factory that constantly improves user experience and increases profitability—attracting more users, further improving the models and profitability. That's a model-driven business.

focus—given Monsanto's deep integration into farms and its data assets—into model-driven farming. Looking to produce more-resilient

making all the other translators more productive in future projects.

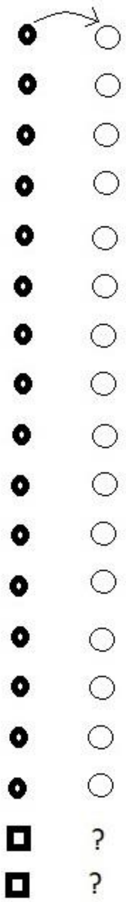
The implications of the rise of model-driven businesses are vast.

because they often have troves of data and startups usually don't. Incumbents will have opportunities to create models with their own data as well as to sell their data to others. Startups will have to be more clever in how they gain access to data and may, in fact, have to acquire incumbents.

Fourth, just as companies have built deep organizational capabilities to manage technology, people, and capital, the same will now happen for models. As the software era took hold, companies everywhere hired chief technology officers, assembled teams of engineers, and designed processes like Agile to deliver software in a systematic, industrialized fashion to their businesses. Companies wanting to become more model-driven will need to create a new discipline of model management—the people, processes and technologies required to develop, validate, deliver and monitor models that create that critical competitive

Chain-Ladder First Link: The Method

- We observe some objects that have changed over time (the circles)
- We observe two new objects (the squares)
- What is an estimate of their changed values?



$$y = bx$$

method

$$b = \frac{\sum y}{\sum x}$$

traditional average link ratio

Let’s play a game

* biased coin idea thanks to https://izbicki.me/blog/how-to-create-an-unfair-coin-and-prove-it-with-math.html

Predictions are not certain: prediction bands

With Excel

Original model:

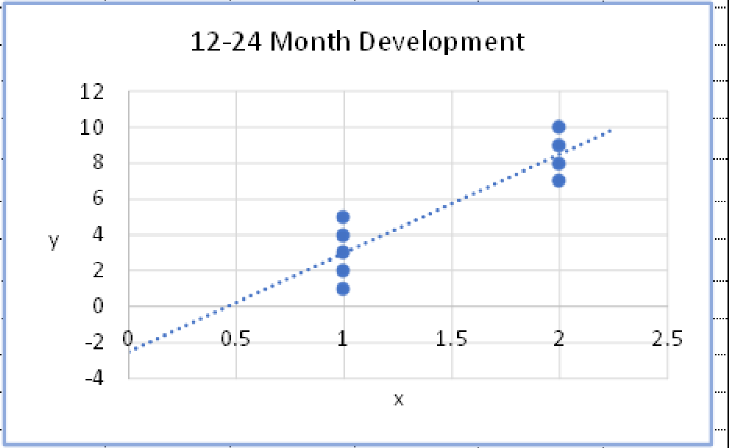
$$y = bx + \sqrt{x}e$$

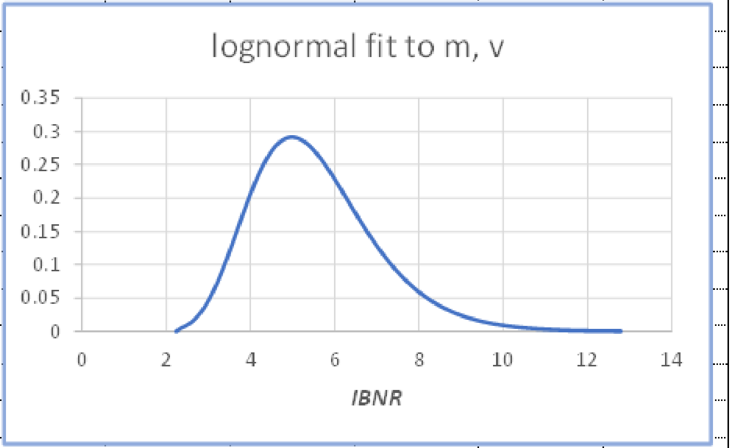
Equivalent model:

$$y' = bx' + e$$

Three stats from equivalent model's data are applied to original model's data

	A	B	C	E	F	G	I	J	K	L
1	Original Data			Equivalent 12-24 month data						
2		x (12 mo)	y (24 mo)		x'=x/√x	y'=y/√x	Model parameters:			
3	1	1	1	1	1.00	1.00	= LINEST(G3:G18,F3:F18,FALSE,TRUE)			
4	2	1	2	2	1.00	2.00	b	3.769	0	const
5	3	1	3	3	1.00	3.00	σ_b	0.40795	#N/A	se_{const}
6	4	1	4	4	1.00	4.00	r_2	0.914	1.471	σ
7	5	1	5	5	1.00	5.00	F	85.4	8	df
8	6	2	7	6	1.41	4.95	ss_{reg}	184.7	17.3	ss_{resid}
9	7	2	8	7	1.41	5.66				
10	8	2	9	8	1.41	6.36				
11	9	2	10	9	1.41	7.07				
12	10	2	7.538	10	1.41	5.33				
13										
14			*IBNR*							
15	Point Estimate						IBNR quantiles:			
16	10	m=	5.538	= 2 * 3.769 - 2			P	log.z	log.tz	log.tz/log.z
17							50%	5.34	5.34	1.00
18	Parameter Risk						60%	5.72	5.73	1.00
19	10		0.816	= 2 * 0.40795			80%	6.68	6.77	1.01
20							90%	7.51	7.74	1.03
21	Process Risk						95%	8.27	8.75	1.06
22	10		2.080	= SQRT(2) * 1.471			99%	9.91	11.53	1.16
23							99.5%	10.59	13.02	1.23
24	Total Risk = Mack S.E.									
25	10	v=	2.234	= SQRT(0.816^2 + 2.080^2)						
26										





* m, v notation c/o Wikipedia (lognormal)

Chain-Ladder First Link: The Model

$$y = bx + \sqrt{x}e$$

error term makes it a model

equivalent model

$$y' = bx' + e$$

where

$$y' = \frac{y}{\sqrt{x}}$$

and

$$x' = \frac{x}{\sqrt{x}}$$

Value of b that minimizes

$$\sum (y' - bx')^2$$

Calculus

is

$$b = \frac{\sum x' y'}{\sum x'^2} = \frac{\sum \frac{x}{\sqrt{x}} \frac{y}{\sqrt{x}}}{\sum \left(\frac{x}{\sqrt{x}}\right)^2} = \frac{\sum y}{\sum x}$$

Chain-Ladder First Link: An Example

- Apply the model

$$y = bx + \sqrt{x}e$$

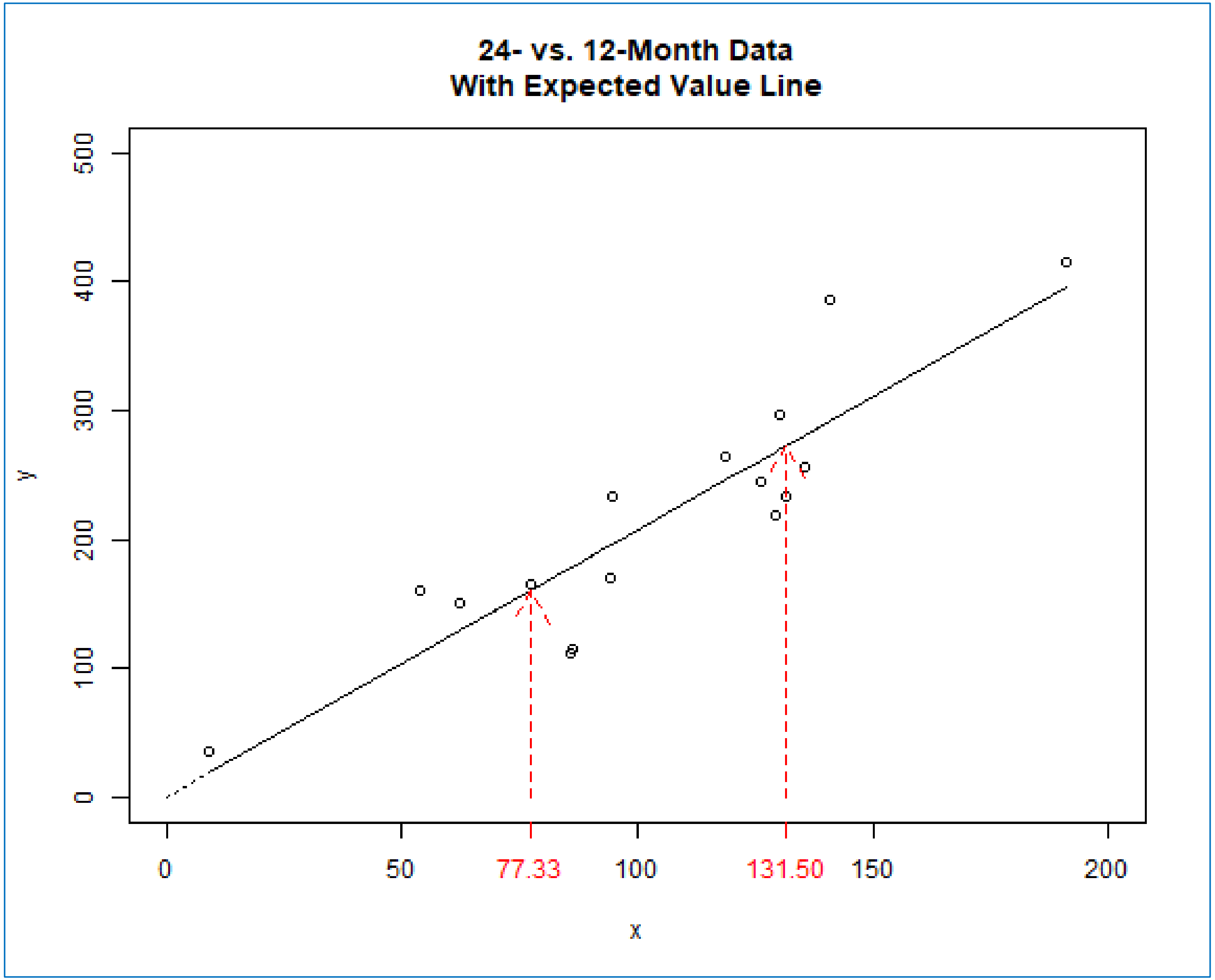
to this skinny triangle

	x	y
1	129.28	218.24
2	135.47	255.51
3	94.53	232.66
4	77.33	165.16
5	130.29	296.19
6	9.10	35.77
7	131.50	233.45
8	86.19	114.70
9	85.79	112.39
10	54.03	161.14
11	94.19	169.68
12	190.87	416.01
13	118.53	263.72
14	126.01	244.73
15	62.47	150.62
16	140.85	385.98
17	77.33	
18	131.50	

12-24 Month Development Experience

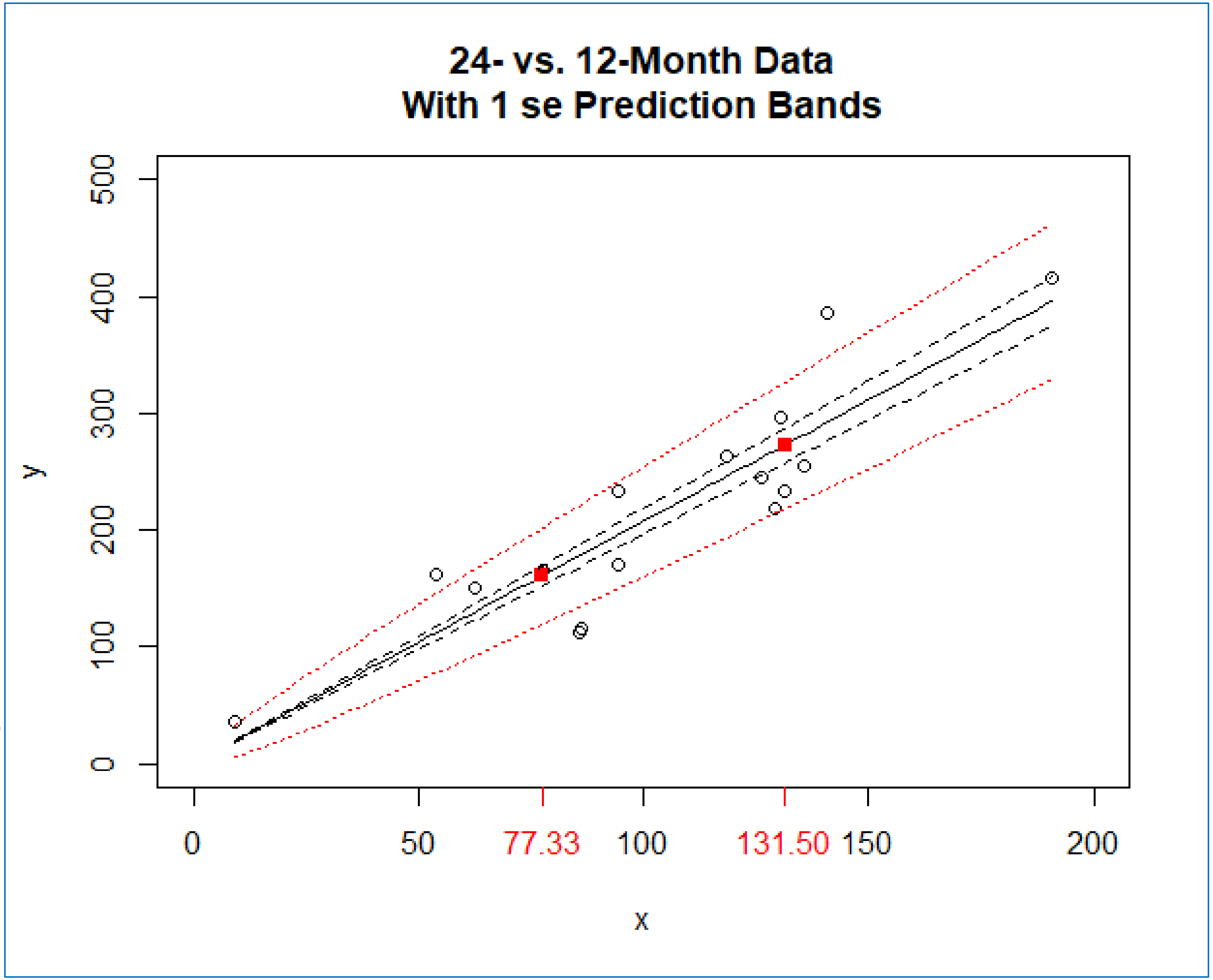
$$b = \frac{\sum y}{\sum x} = 2.074$$

- 2.074 = slope of the line through origin
- prediction of new initial observations:
  77.33 -> 160.4
  131.5 -> 272.7



Predictions are not certain: prediction bands

- --- Parameter risk Δ Variability of estimated mean
- Process risk Γ Variability around theoretical mean
- ...... Total risk=$\sqrt{\Delta^2 + \Gamma^2}$ Variability of a predicted outcome

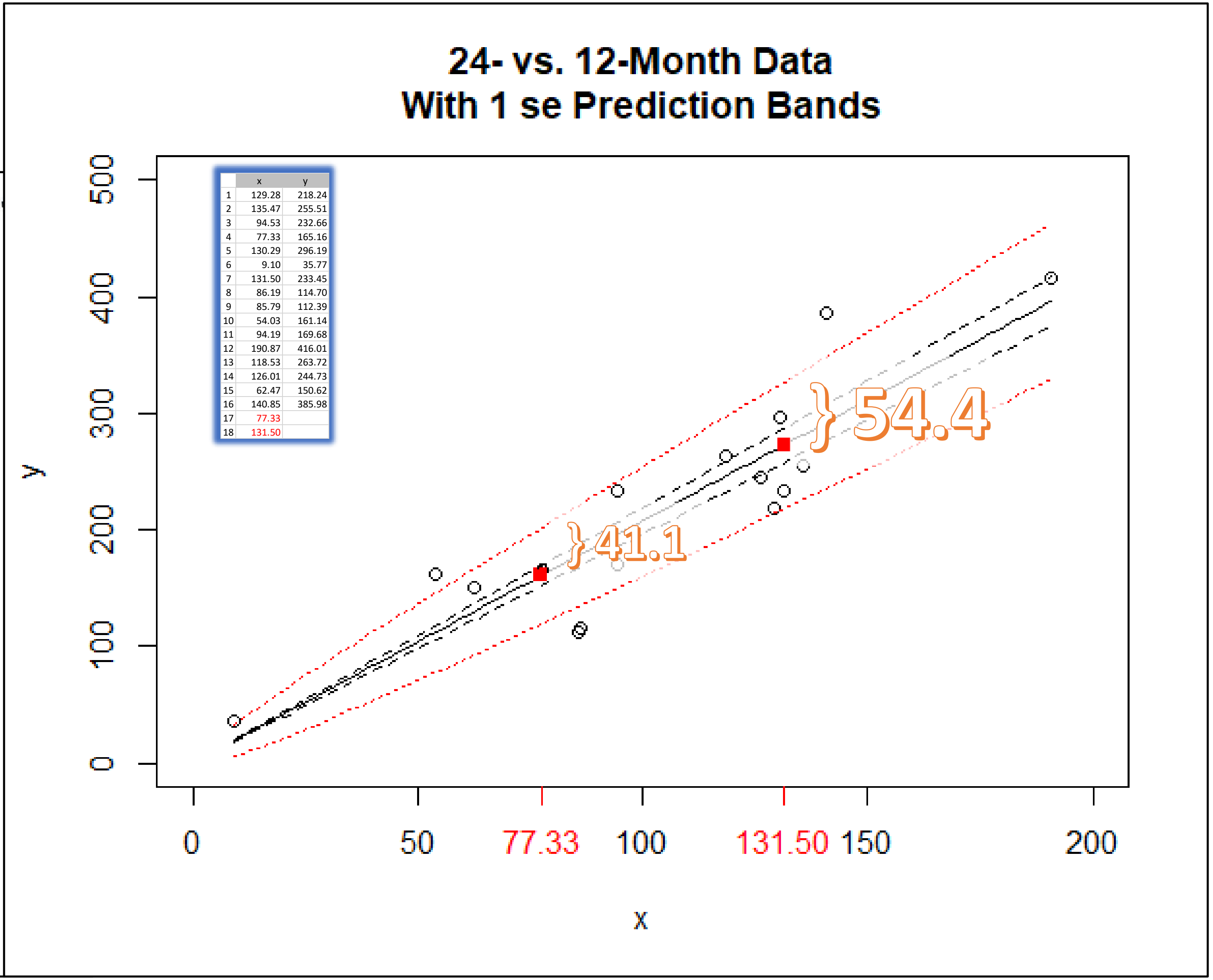


* notation by Ali Majidi

Predictions are not certain: prediction bands

With R

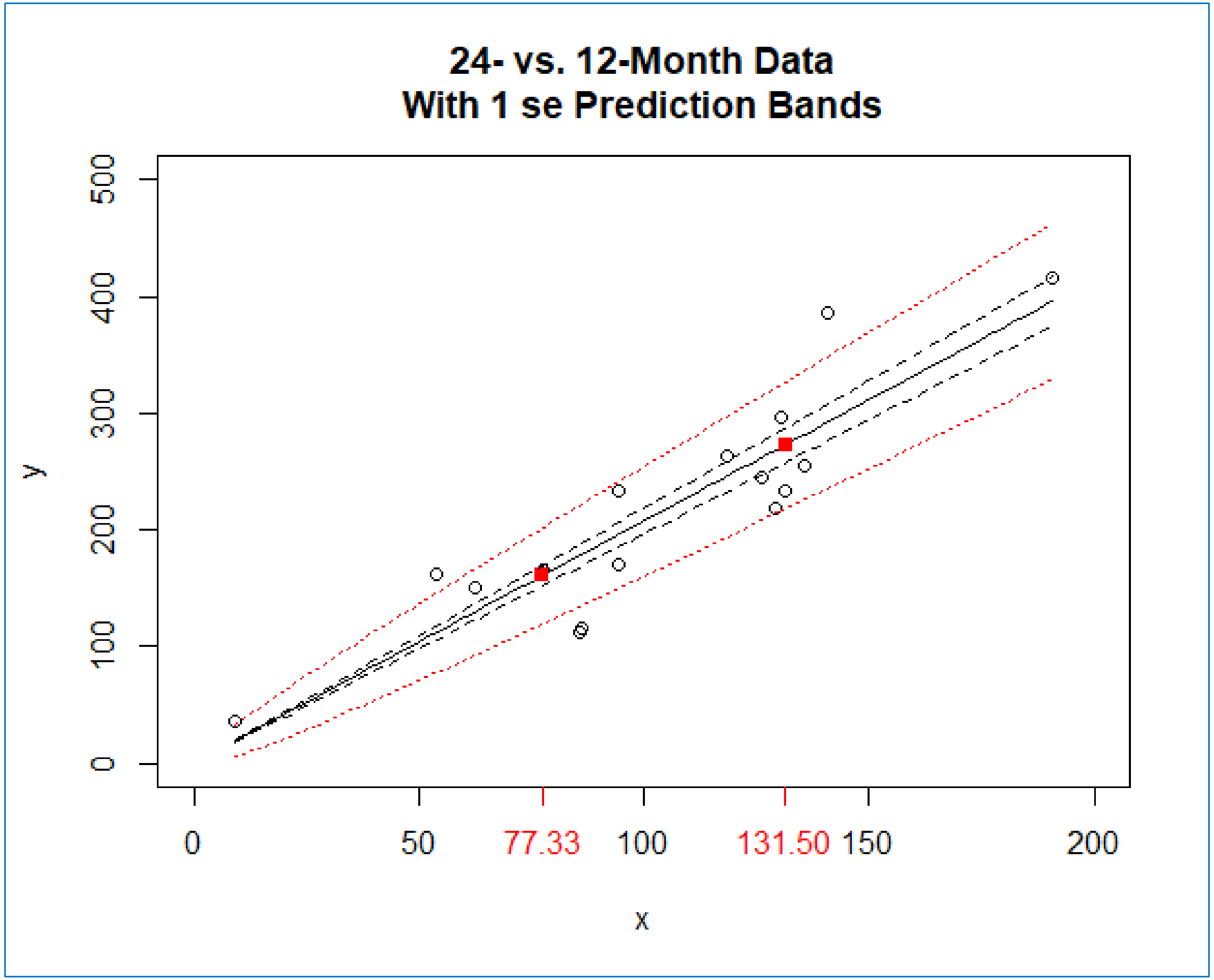
```
> ChainLadder::MackChainLadder(tri, est.sigma = "Mack")
ChainLadder::MackChainLadder(Triangle = tri, est.sigma = "

   Latest Dev.To.Date Ultimate IBNR Mack.S.E CV(IBNR)
1   218.2       1.000    218.2    0      0.0      NaN
2   255.5       1.000    255.5    0      0.0      NaN
3   232.7       1.000    232.7    0      0.0      NaN
4   165.2       1.000    165.2    0      0.0      NaN
5   296.2       1.000    296.2    0      0.0      NaN
6    35.8       1.000     35.8    0      0.0      NaN
7   233.4       1.000    233.4    0      0.0      NaN
8   114.7       1.000    114.7    0      0.0      NaN
9   112.4       1.000    112.4    0      0.0      NaN
10  161.1       1.000    161.1    0      0.0      NaN
11  169.7       1.000    169.7    0      0.0      NaN
12  416.0       1.000    416.0    0      0.0      NaN
13  263.7       1.000    263.7    0      0.0      NaN
14  244.7       1.000    244.7    0      0.0      NaN
15  150.6       1.000    150.6    0      0.0      NaN
16  386.0       1.000    386.0    0      0.0      NaN
17   77.3       0.482    160.4   83     41.1    0.495
18  131.5       0.482    272.7  141     54.4    0.385

              Totals
Latest:    3,664.78
Dev:           0.94
Ultimate:  3,889.04
IBNR:        224.26
Mack.S.E      70.00
CV(IBNR):      0.31
```



	x	y
1	129.28	218.24
2	135.47	255.51
3	94.53	232.66
4	77.33	165.16
5	130.29	296.19
6	9.10	35.77
7	131.50	233.45
8	86.19	114.70
9	85.79	112.39
10	54.03	161.14
11	94.19	169.68
12	190.87	416.01
13	118.53	263.72
14	126.01	244.73
15	62.47	150.62
16	140.85	385.98
17	77.33	
18	131.50	

* ChainLadder package by Markus Gesmann et.al.

Why does the prediction envelope fan out only at the high end?



Chain-Ladder link

$$y = bx + \sqrt{x}e$$

hint

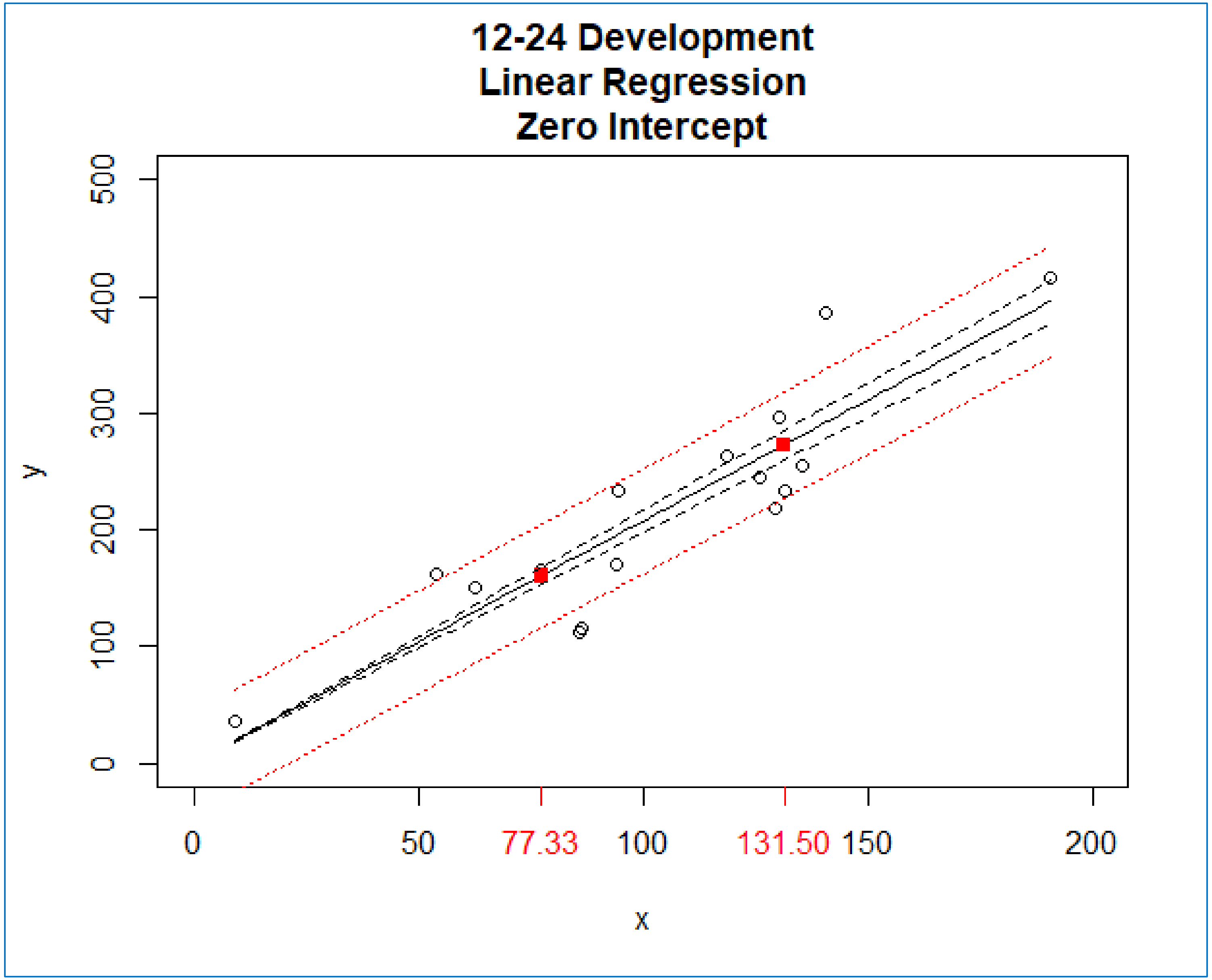
- Assumption is, The higher the initial value, the greater the variability of the subsequent value
- When might you have less variability the larger the beginning value?

How do prediction bands look under different models?

Prediction bands without square-root-o-skedasticity

$$y = bx + e$$

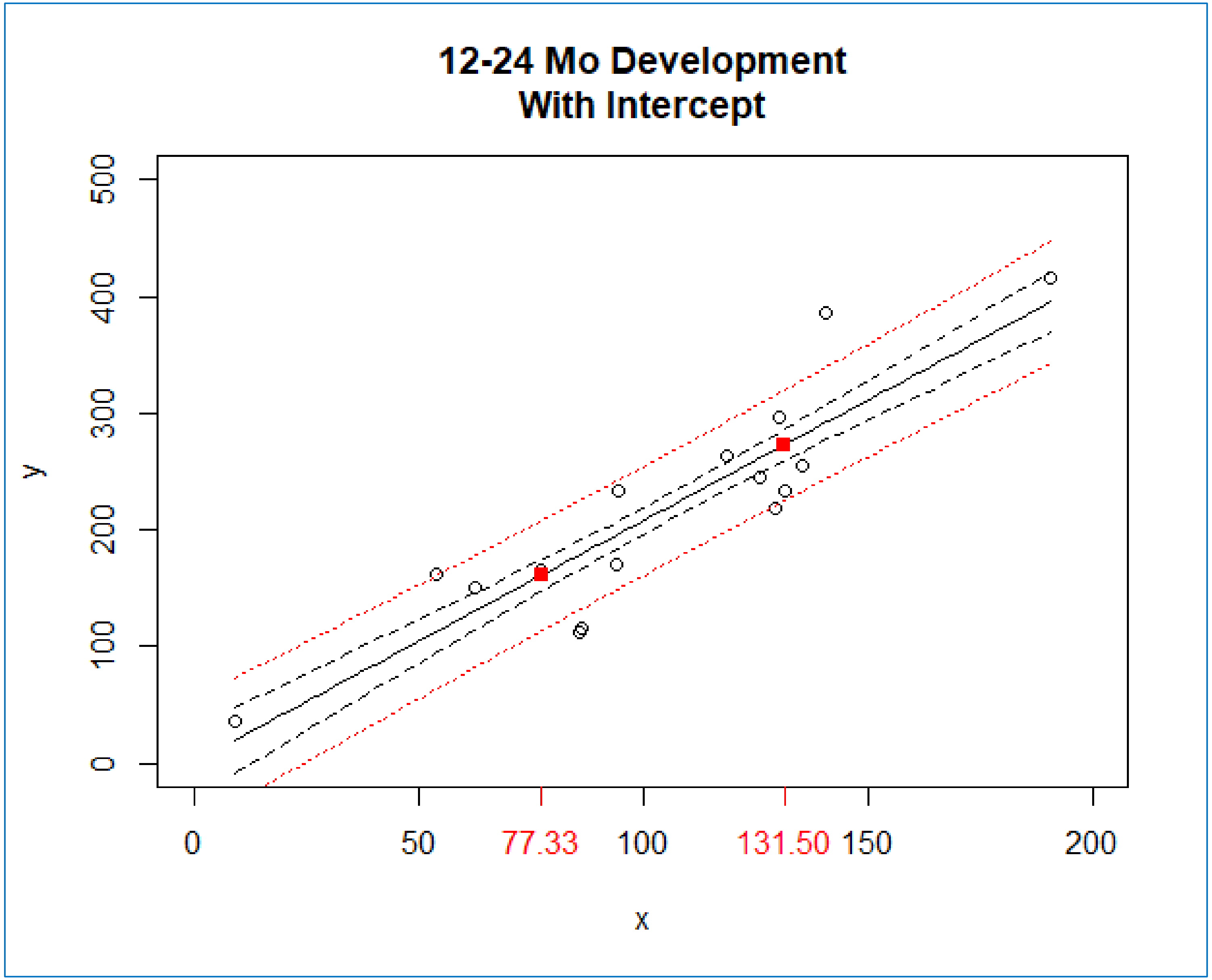
- --- parameter risk
- ······ total risk



Prediction bands when there's an intercept

$$y = a + bx + \ e$$

- --- parameter risk
- ⋯⋯ total risk



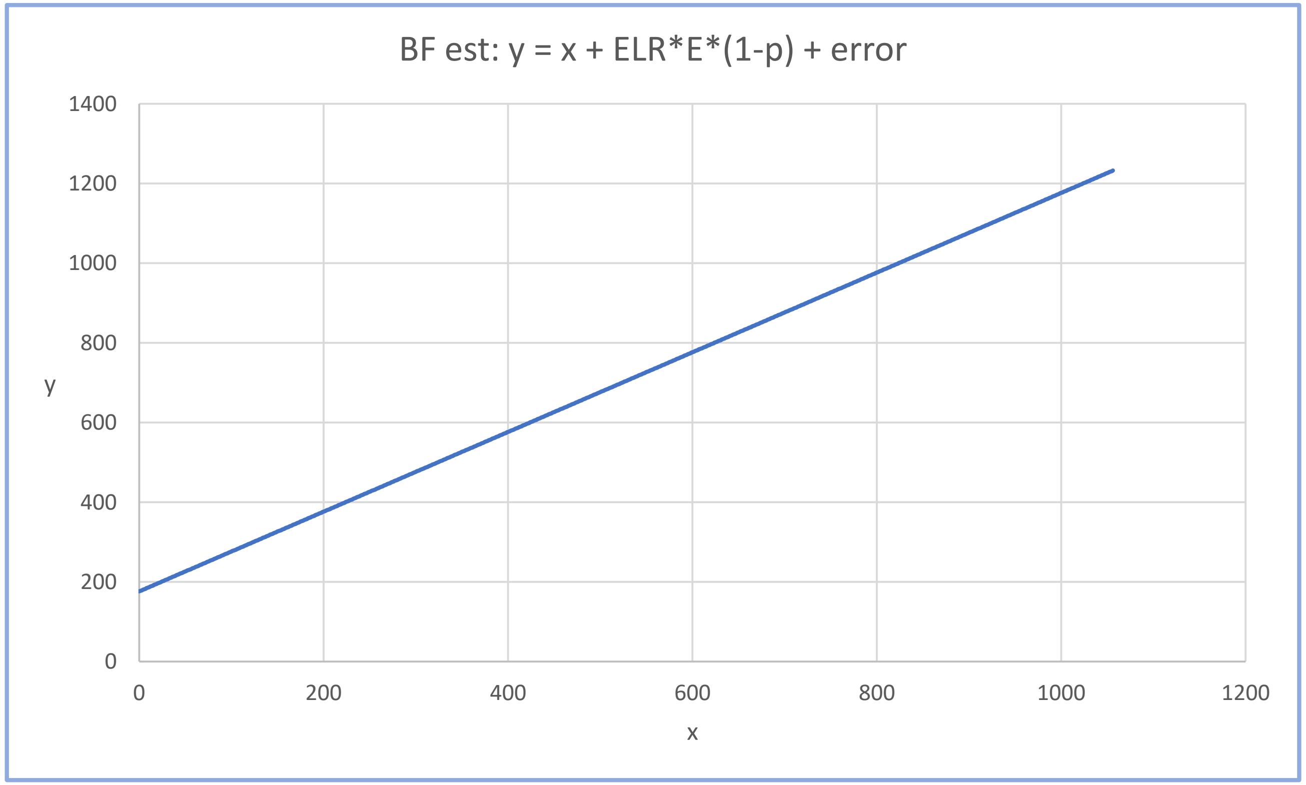
Homework

1. What would the graph of the model look like if the simple average is the optimal link ratio?
   - Does the answer change if "optimal" is a matter of actuarial judgment?
2. What could be drivers of a non-zero intercept?
3. How to model the BF method within the Chain-Ladder paradigm?
4. How to model the first column within the Chain-Ladder paradigm?
5. Prove that our game satisfies the assumptions of the model

$$y = bx + \sqrt{x}e$$

Bornhuetter-Ferguson

- What is the slope of the line?
- What is the intercept?

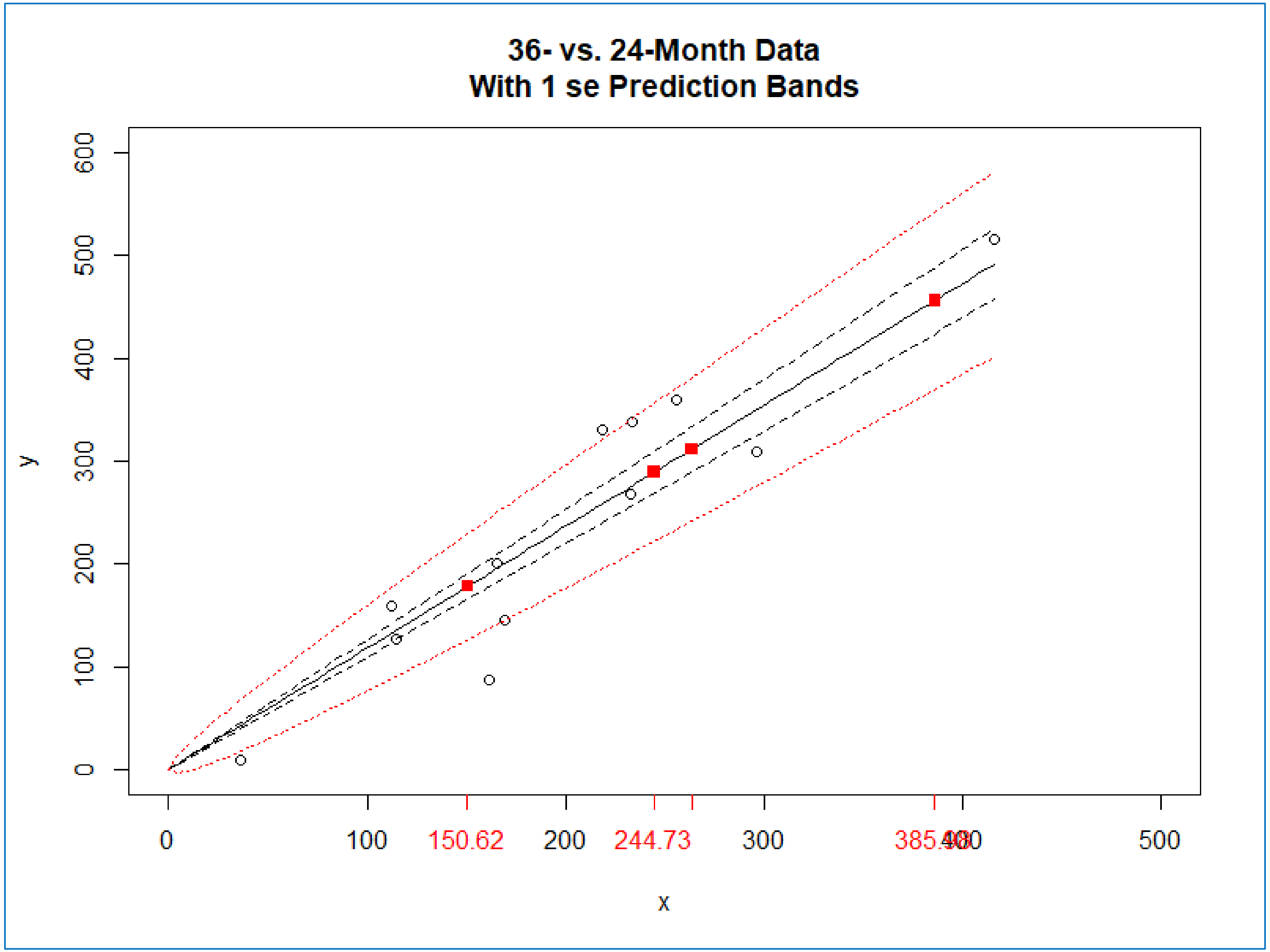


Chain-Ladder Second Link: Add Another Column

	x (12 mo)	y (24 mo)	z (36 mo)
1	129.28	218.24	330.88
2	135.47	255.51	359.34
3	94.53	232.66	267.56
4	77.33	165.16	200.61
5	130.29	296.19	309.08
6	9.10	35.77	9.53
7	131.50	233.45	337.82
8	86.19	114.70	127.00
9	85.79	112.39	159.52
10	54.03	161.14	86.60
11	94.19	169.68	145.21
12	190.87	416.01	514.95
13	118.53	263.72	
14	126.01	244.73	
15	62.47	150.62	
16	140.84	385.98	
17	77.33		
18	131.50		

Chain-Ladder Second Link: Add Another Column

- $b_y = 1.181$
- $sigma\_b_y = 0.083$
- $sigma_y = 4.1$



Chain-Ladder predicts the future recursively

- Orange projections are products of a scalar and an estimated parameter **Which is which?**
  - Formulas for Parameter Risk and Process Risk can be found in slides above
- Red projections are products of an estimate and an estimated parameter
  - Formulas for Parameter Risk and Process Risk are derived from the

Law of Total Variance

	x (12 mo)	y (24 mo)	z (36 mo)
1	129.28	218.24	330.88
2	135.47	255.51	359.34
3	94.53	232.66	267.56
4	77.33	165.16	200.61
5	130.29	296.19	309.08
6	9.10	35.77	9.53
7	131.50	233.45	337.82
8	86.19	114.70	127.00
9	85.79	112.39	159.52
10	54.03	161.14	86.60
11	94.19	169.68	145.21
12	190.87	416.01	514.95
13	118.53	263.72	**311.45**
14	126.01	244.73	**289.03**
15	62.47	150.62	**177.88**
16	140.84	385.98	**455.84**
17	77.33	**160.38**	**189.41**
18	131.50	**272.73**	**322.10**
b	**2.074**	**1.181**	

Law of Total Variance

sidebar

- Wikipedia:

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

- "In actuarial science, specifically credibility theory, the first component is called the expected value of the process variance (**EVPV**) and the second is called the variance of the hypothetical means (**VHM**)."
  - Retrieved June 25, 2015

- See Majidi and Bardis formula derivations, "A Family of Chain-Ladder Models," *Variance*, Vol 6, Issue 2, pp. 157-158

Recursive projections with statistics – complete many squares

- Expected Value
- Parameter Risk – Δ

$$\Delta_y^2 = x^2 \cdot \widehat{\sigma_b}^2$$
$$\Delta_z^2 = \hat{y}^2 \cdot \widehat{\sigma_b}^2 + \hat{b}^2 \cdot \Delta_y^2 + \widehat{\sigma_b}^2 \cdot \Delta_y^2$$

- Process Risk – Γ

$$\Gamma_y^2 = x \cdot \widehat{\sigma_x}^2$$
$$\Gamma_z^2 = \hat{y} \cdot \widehat{\sigma_y}^2 + \hat{b}^2 \cdot \Gamma_y^2$$

Data

	x	y	z
1	xxx	yyy	zzz
2	xxx	yyy	
3	xxx		

Parameter estimates

b_x	b_y
σ_b	σ_b
σ_x	σ_y

Expected Values

	x	y	z
1			
2			z^ = yyy*b_y
3		y^ = xxx*b_x	z^ = y^*b_y

Parameter Risk

	x	y	z
1			
2			Δ_z
3		Δ_y	Δ_z

Process Risk

	x	y	z
1			
2			Γ_z
3		Γ_y	Γ_z

Tricks for Risk Estimates for the Total/sum Row

Expected Values

	x	y	z
1			
2		yyy	$z\hat{} = yyy * b_y$
3		$y\hat{} = xxx * b_x$	$z\hat{} = y\hat{} * b_y$
future sum		$Y\hat{}$	$Z\hat{} = (yyy + Y\hat{}) * b_y$

Parameter Risk

	x	y	z
1			
2			Δ_z
3		Δ_y	Δ_z
future sum		Δ_y	same formula

Process Risk

	x	y	z
1			
2			Γ_z
3		Γ_y	Γ_z
future sum		Γ_y	$\Gamma_Z^2 = \Gamma_z^2 + \Gamma_z^2$

Why not directly estimate the 12-36 month link ratio?

- What if you learned b_{xz} = sum(z) / sum(x) = 2.337
- Why not say the expected 36-month value of x = 77.33 is

  77.33 * 2.337 = 180.7 (se 44.7)

  rather than

  77.33 * 2.074 * 1.181 = 189.4 (se 72.1 see above)

Can you ignore y or not?

- There is important information in the 24-month value
- The path to ultimate is important
- *It's the journey*

	x (12 mo)	y (24 mo)	z (36 mo)	
1	129.28	218.24	330.88	
2	135.47	255.51	359.34	
3	94.53	232.66	267.56	
4	77.33	165.16	200.61	
5	130.29	296.19	309.08	
6	9.10	35.77	9.53	
7	131.50	233.45	337.82	
8	86.19	114.70	127.00	
9	85.79	112.39	159.52	
10	54.03	161.14	86.60	
11	94.19	169.68	145.21	
12	190.87	416.01	514.95	***IBNR***
13	118.53	263.72	277.03	13.31
14	126.01	244.73	294.52	49.79
15	62.47	150.62	146.01	-4.61
16	140.84	385.98	329.18	-56.80
17	77.33		180.74	103.41
18	131.50		307.35	175.85
b	2.337			

Continue the Journey – Chain to Ultimate

- Recursive estimates are carried forward to the last pair of development columns
- Technical considerations
  - What to do when there are not enough observations to get a good estimate of sigma (zero degrees of freedom)
  - What to do with a tail
- Mack has recommendations for handling these technicalities
- The ChainLadder package's MackChainLadder function includes Mack's recommendations, as well as others
- **Let's see some examples**

California WCIRB Agenda June 2018
Combined Indemnity and Medical Incurred

($M)	15	27	39	51	63	75	87	99	111	123	135	147	159	171	183	195	207	219	231	243	255	267	279	291	303	315	327	339	351	363	375	389
1986																						2,510	2,513	2,521	2,528	2,537	2,545	2,551	2,554	2,555	2,552	2,557
1987																					2,834	2,841	2,854	2,853	2,864	2,870	2,877	2,881	2,884	2,882	2,885	
1988																				3,055	3,066	3,073	3,084	3,089	3,098	3,101	3,106	3,109	3,111	3,108		
1989																			3,568	3,577	3,593	3,604	3,617	3,629	3,629	3,633	3,633	3,633	3,633			
1990																		1,943	1,945	1,945	1,945	1,947	1,947	4,041	4,034	4,035	4,037	4,037				
1991																	4,671	4,688	4,704	4,715	4,720	4,729	4,733	4,740	4,742	4,740	4,738					
1992																3,739	3,749	3,765	3,769	3,776	3,787	3,794	3,800	3,798	3,796	3,800						
1993															2,993	3,015	3,026	3,034	3,055	3,077	3,080	3,083	3,082	3,075	3,078							
1994														3,093	3,103	3,133	3,151	3,169	3,180	3,191	3,199	3,202	3,197	3,193								
1995													3,240	3,288	3,316	3,328	3,346	3,370	3,361	3,372	3,373	3,368	3,366									
1996												3,645	3,683	3,709	3,729	3,754	3,771	3,787	3,795	3,795	3,798	3,797										
1997											4,333	4,387	4,428	4,454	4,480	4,493	4,504	4,504	4,493	4,487	4,483											
1998										5,359	5,419	5,478	5,516	5,554	5,590	5,618	5,655	5,658	5,663	5,655												
1999									5,923	6,025	6,091	6,149	6,205	6,242	6,286	6,302	6,302	6,296	6,286													
2000								6,768	6,867	6,950	7,041	7,113	7,185	7,237	7,267	7,266	7,247	7,246														
2001							9,536	9,791	10,001	10,196	10,338	10,461	10,586	10,635	10,628	10,632	10,614															
2002						9,442	9,715	9,935	10,134	10,323	10,463	10,561	10,611	10,604	10,613	10,617																
2003					8,439	8,782	9,054	9,329	9,576	9,758	9,893	9,977	9,992	10,000	10,014																	
2004				5,996	6,364	6,665	6,971	7,196	7,372	7,529	7,586	7,623	7,628	7,623																		
2005			4,606	4,976	5,338	5,678	5,959	6,154	6,293	6,386	6,418	6,445	6,470																			
2006		4,158	4,759	5,221	5,609	5,928	6,174	6,360	6,459	6,538	6,570	6,591																				
2007	2,999	4,296	5,109	5,670	6,085	6,445	6,681	6,852	6,947	7,009	7,039																					
2008	2,996	4,391	5,319	5,936	6,394	6,726	6,934	7,067	7,136	7,189																						
2009	2,823	4,283	5,150	5,816	6,250	6,542	6,695	6,794	6,858																							
2010	2,919	4,421	5,425	6,048	6,445	6,680	6,833	6,925																								
2011	2,921	4,464	5,362	5,928	6,263	6,437	6,543																									
2012	3,124	4,654	5,490	5,952	6,276	6,454																										
2013	3,358	4,859	5,602	6,072	6,309																											
2014	3,448	4,963	5,823	6,282																												
2015	3,695	5,306	6,113																													
2016	3,824	5,392																														
2017	3,350																															

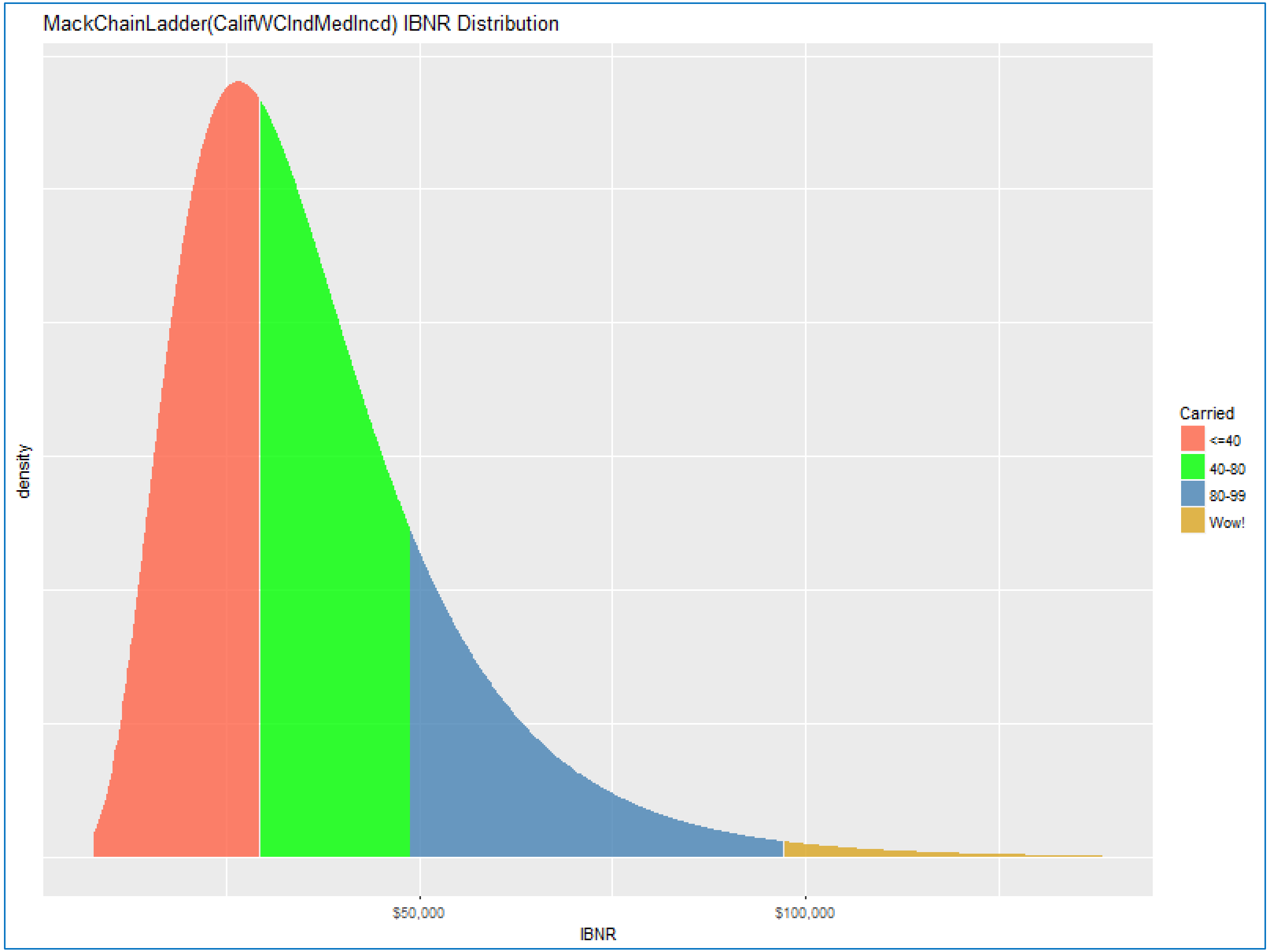
* triangle creation approach thanks to Dave Bellusci; data entry thanks to Connan Houser

MackChainLadder(WCIRB Indemnity + Medical Combined Incurred, tail = 1.025)

	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1986	2,557	0.976	2,621	64	35	55%
1987	2,885	0.974	2,963	78	38	49%
1988	3,108	0.974	3,192	84	40	47%
1989	3,633	0.974	3,729	96	43	45%
1990	4,037	0.974	4,147	110	46	42%
1991	4,738	0.973	4,871	133	50	38%
1992	3,800	0.972	3,910	110	45	41%
1993	3,078	0.971	3,170	92	40	44%
1994	3,193	0.970	3,292	99	42	42%
1995	3,366	0.904	3,724	358	1,024	286%
1996	3,797	0.903	4,207	410	1,096	268%
1997	4,483	0.901	4,973	490	1,204	246%
1998	5,655	0.900	6,283	628	1,378	219%
1999	6,286	0.899	6,994	708	1,467	207%
2000	7,246	0.898	8,070	824	1,598	194%
2001	10,614	0.896	11,846	1,232	2,025	164%
2002	10,617	0.895	11,869	1,252	2,028	162%
2003	10,014	0.892	11,224	1,210	1,958	162%
2004	7,623	0.889	8,572	949	1,657	175%
2005	6,470	0.886	7,304	834	1,506	181%
2006	6,591	0.880	7,489	898	1,529	170%
2007	7,039	0.873	8,066	1,027	1,599	156%
2008	7,189	0.864	8,323	1,134	1,630	144%
2009	6,858	0.850	8,064	1,206	1,599	133%
2010	6,925	0.835	8,291	1,366	1,627	119%
2011	6,543	0.815	8,029	1,486	1,597	108%
2012	6,454	0.790	8,175	1,721	1,617	94%
2013	6,309	0.755	8,353	2,044	1,641	80%
2014	6,282	0.710	8,854	2,572	1,703	66%
2015	6,113	0.645	9,476	3,363	1,782	53%
2016	5,392	0.546	9,883	4,491	1,842	41%
2017	3,350	0.372	8,994	5,644	1,759	31%
sum	182,245		218,958	36,713	18,023	49%

Reported industry IBNR @ 3/31/2018
= $36,196
~58%-ile

Very close to MackChainLadder central estimate
= $36,713



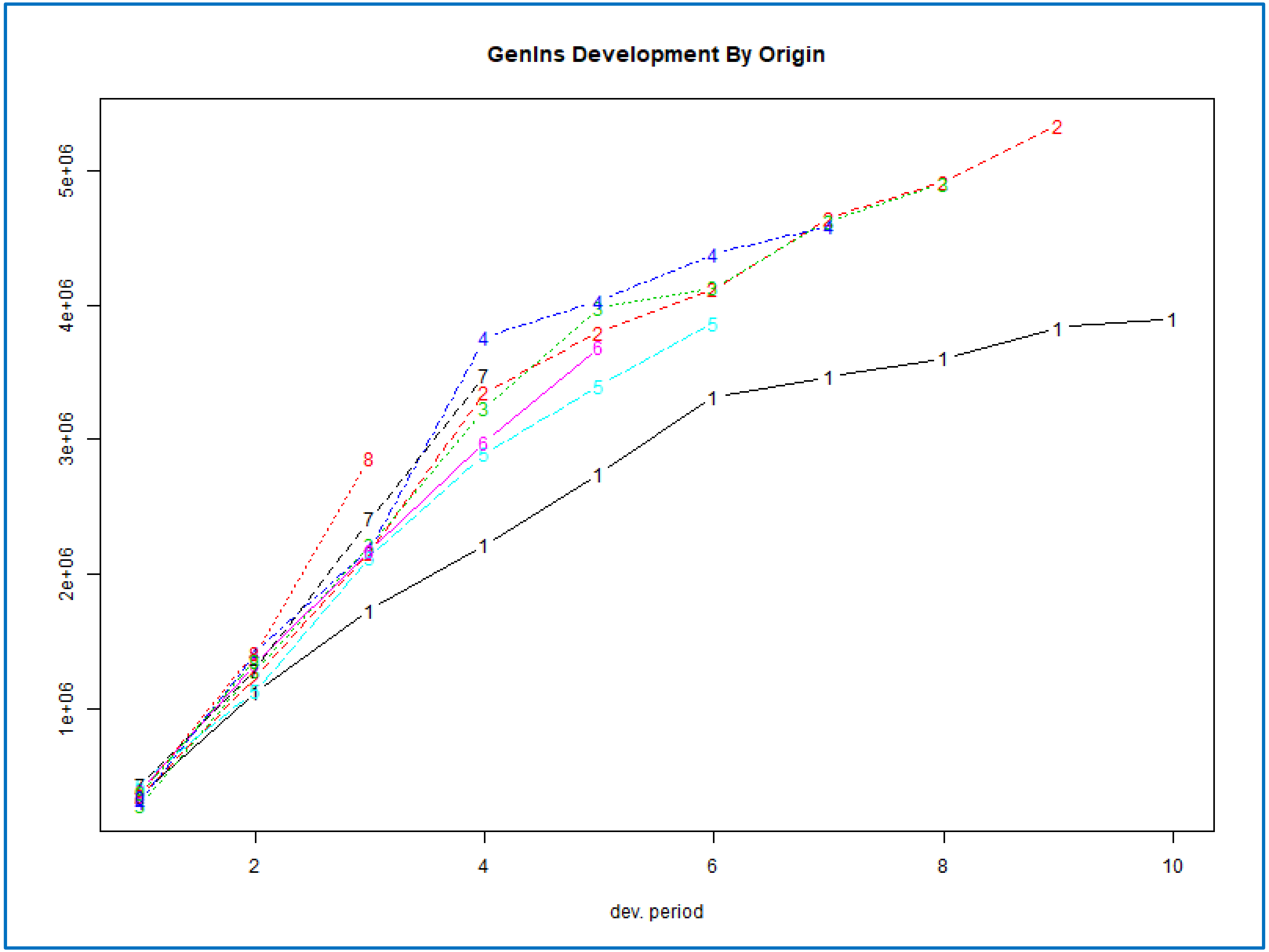
ChainLadder sample GL triangle 'GenIns' (in thousands)

origin	1	2	3	4	5	6	7	8	9	10
1	358	1,125	1,735	2,218	2,746	3,320	3,466	3,606	3,834	3,901
2	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339	
3	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909		
4	311	1,419	2,195	3,757	4,030	4,382	4,588			
5	443	1,136	2,128	2,898	3,403	3,873				
6	396	1,333	2,181	2,986	3,692					
7	441	1,288	2,420	3,483						
8	359	1,421	2,864							
9	377	1,363								
10	344									

origin	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
1	3.143	1.543	1.278	1.238	1.209	1.044	1.04	1.063	1.018
2	3.511	1.755	1.545	1.133	1.084	1.128	1.057	1.086	
3	4.448	1.717	1.458	1.232	1.037	1.12	1.061		
4	4.568	1.547	1.712	1.073	1.087	1.047			
5	2.564	1.873	1.362	1.174	1.138				
6	3.366	1.636	1.369	1.236					
7	2.923	1.878	1.439						
8	3.953	2.016							
9	3.619								

GenIns is a triangle first published in Taylor & Ashe paper (1983) and repeatedly studied in the literature

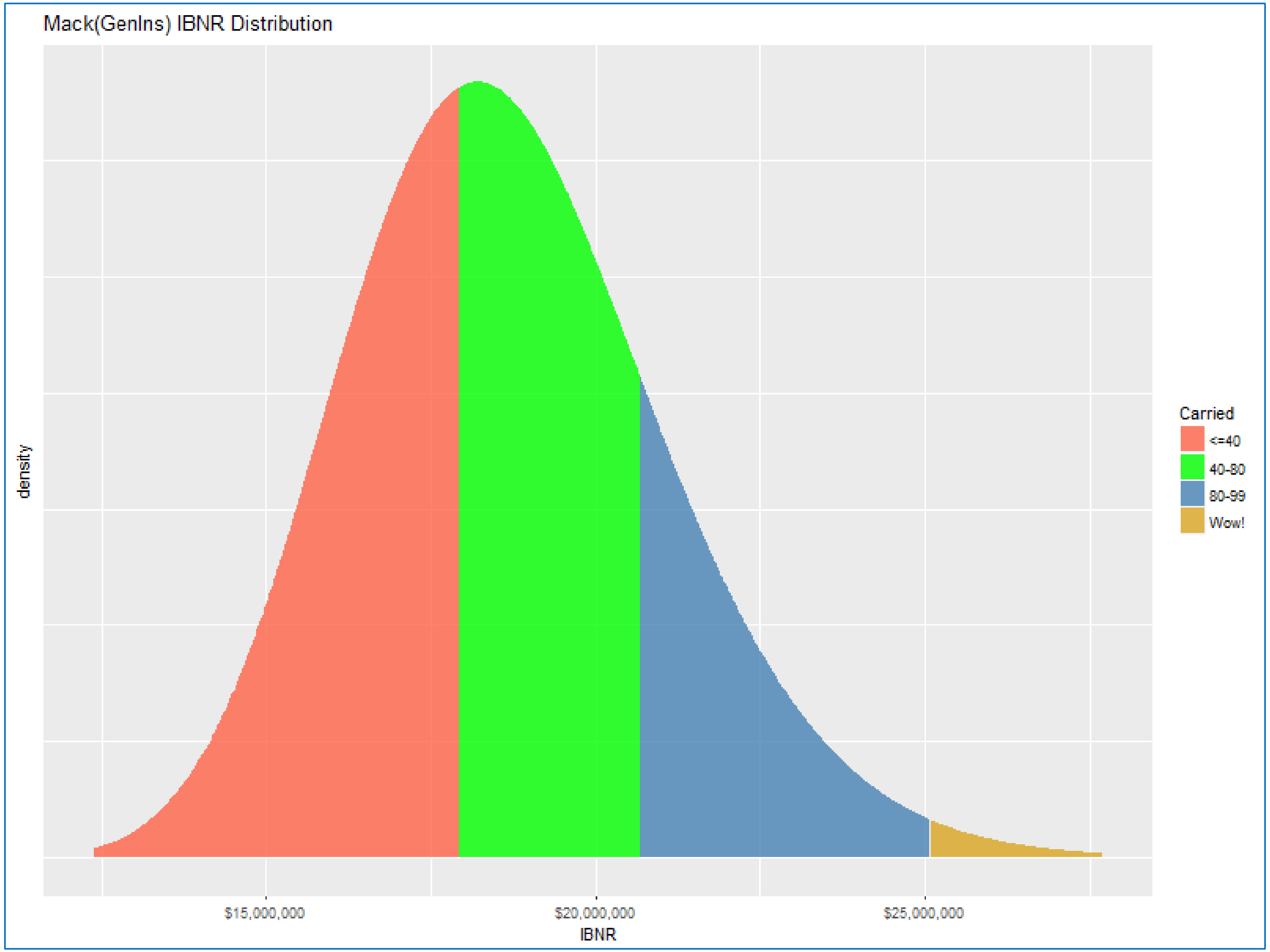
ChainLadder:: plot(GenIns)



MackChainLadder(GenIns)

origin	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1	3,901	100.0%	3,901	0	0	
2	5,339	98.3%	5,434	95	72	75.9%
3	4,909	91.3%	5,379	470	119	25.4%
4	4,588	86.6%	5,298	710	132	18.5%
5	3,873	79.7%	4,858	985	261	26.5%
6	3,692	72.2%	5,111	1,419	410	28.9%
7	3,483	61.5%	5,661	2,178	558	25.6%
8	2,864	42.2%	6,785	3,920	875	22.3%
9	1,363	24.2%	5,642	4,279	971	22.7%
10	344	6.9%	4,970	4,626	1,363	29.5%
sum	34,358		53,039	18,681	2,441	13.1%

Safety Levels of GenIns Carried IBNR



- Wrap-up: What are possible uses of an IBNR distribution?
- Rather than a distribution, can Mack/Murphy be used in predictive analytics?

What happens when Mack/Murphy is run on detail data?

- Suppose x and y are actually observations from 4 companies in 4 accident years
- Will link ratios from aggregated data and detail data always be the same?
- What about the risk statistics?

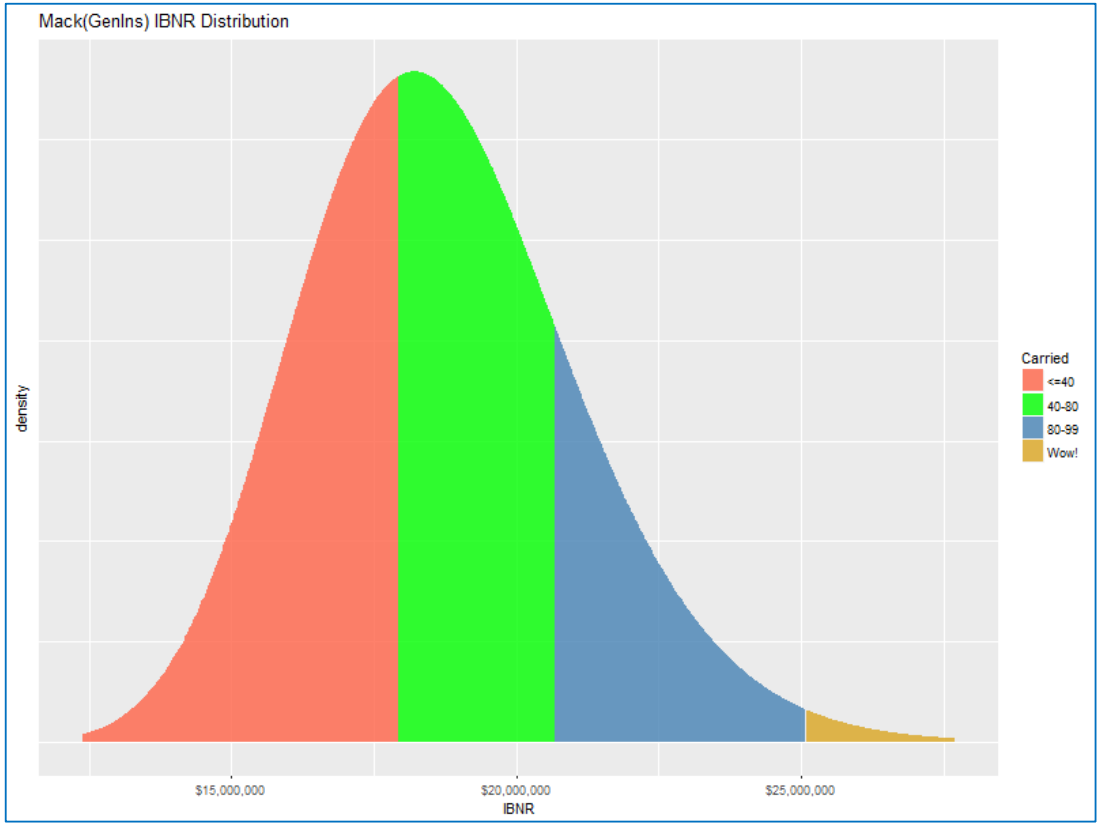
	x	y	x	y
1	129.28	218.24	436.61	871.57
2	135.47	255.51		
3	94.53	232.66		
4	77.33	165.16		
5	130.29	296.19	357.09	680.11
6	9.10	35.77		
7	131.50	233.45		
8	86.19	114.70		
9	85.79	112.39	424.88	859.22
10	54.03	161.14		
11	94.19	169.68		
12	190.87	416.01		
13	118.53	263.72	447.86	1045.05
14	126.01	244.73		
15	62.47	150.62		
16	140.85	385.98		
17	77.33			
18	131.50			
	1666.44	3455.95	1666.44	3455.95
	b_{detail} =	2.074	$b_{aggregated}$ =	2.074

GenIns at the Claim Level

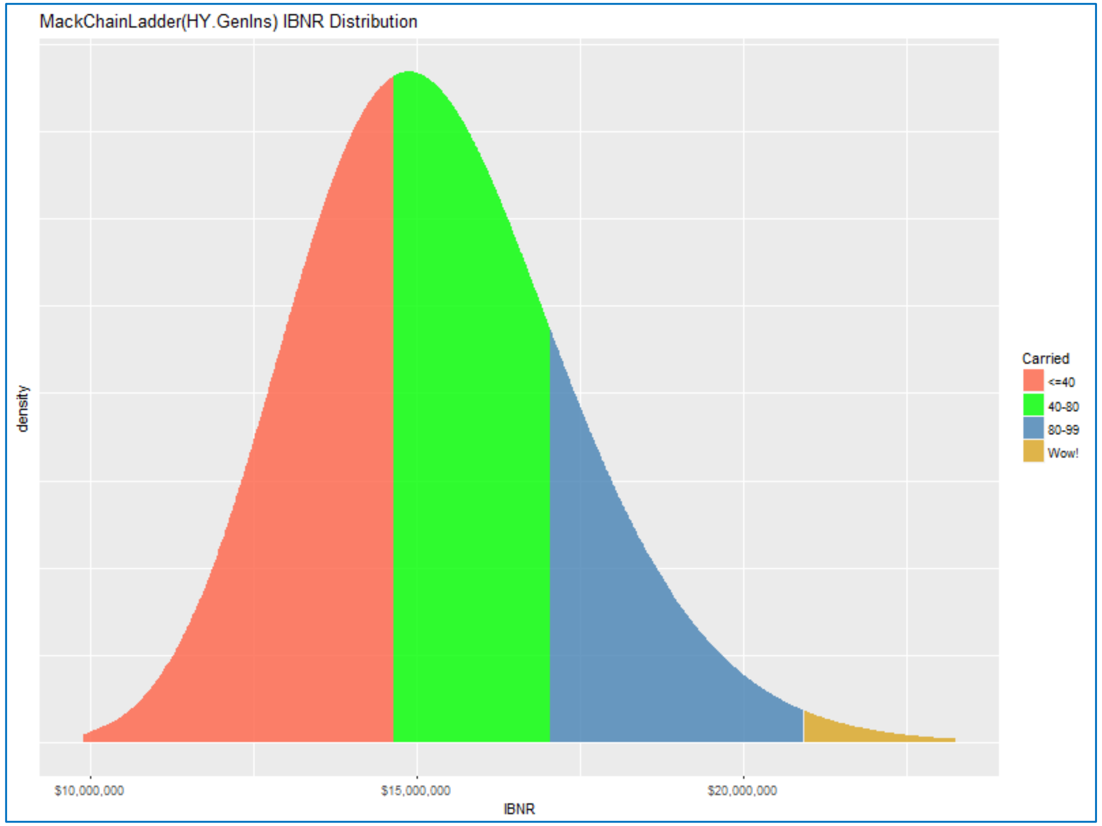
- Hai You generated simulations of over 6000 synthetic claims whose accident year aggregation is “close in shape” to GenIns
  - We pegged the 13% cv as the primary measurement of similarity
- Hai’s claim-characteristic choices included:
  - Frequency distribution
  - Severity distribution
  - Distribution for the number of payments per claim
  - Report lag and payment lag
- The purpose of this exercise was to compare the Mack results on the claim detail with the statistics from the aggregate triangle

IBNR distributions from aggregated triangles are very similar

Original GenIns



Aggregated triangle from Hai data



GenIns at the Claim Level:
Claim detail sample in triangle format

	Latest	Ultimate	IBNR	Mack.S.E	CV(IBNR)
GenIns	34,358	53,039	18,681	2,441	13.1%
HY.GenIns	32,556	47,866	15,310	2,127	13.9%
HY detail	32,556	40,909	8,353	695	8.3%

3304	
3305	
4330	9,102
5538	3,144
5911	1,007
6289	273
6300	1,425

Why the IBNR drop?

Why the CV drop?

* simulated claims by Hai You

Is the weighted average development factor appropriate?

Dev Factor	12-24
GenIns	3.491
HY.GenIns	3.413
HY.detail	1.288

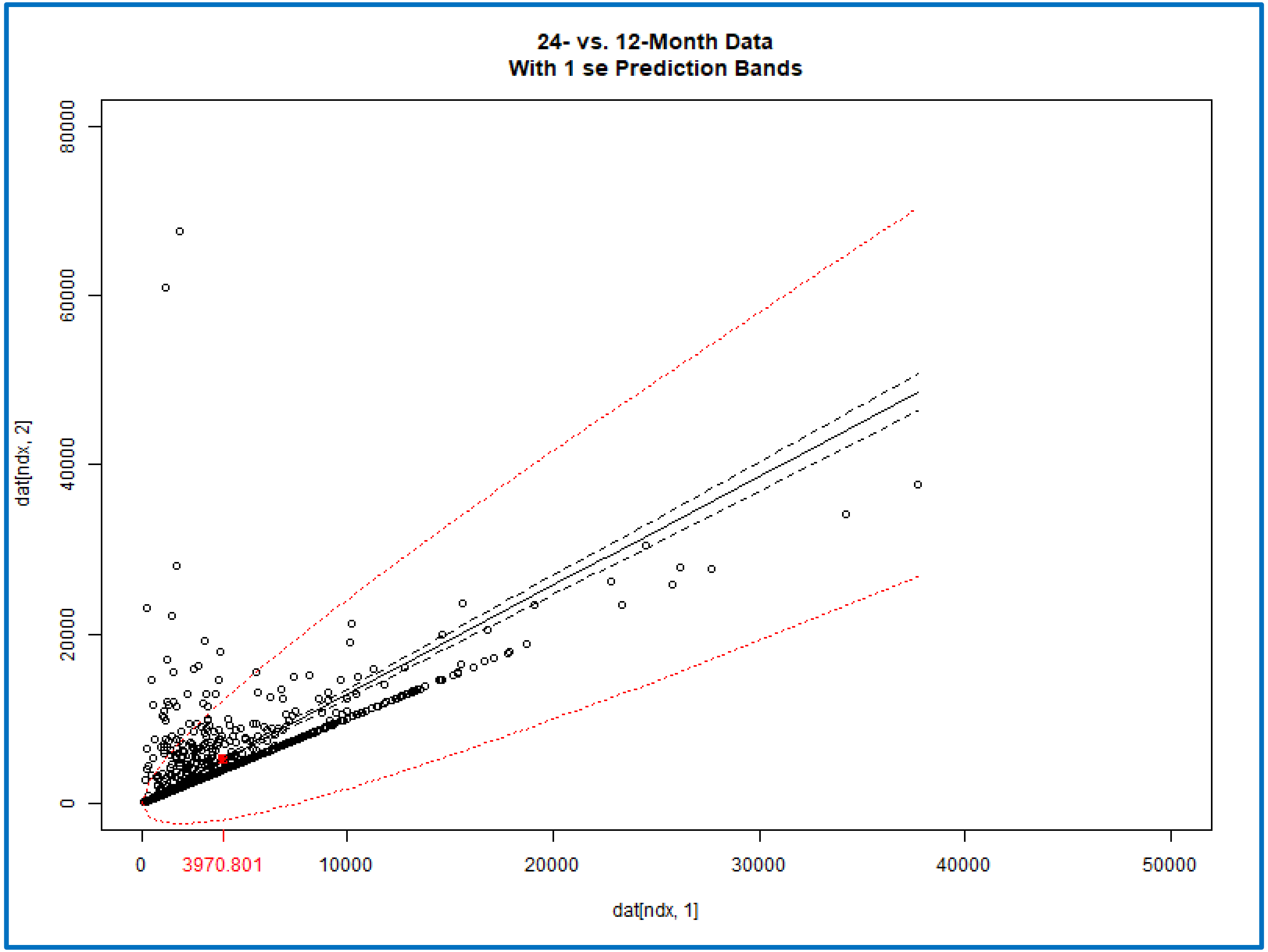
What happened to the 12-24 factor from the claim detail?!?

The 12-24 month relationship from the claim detail

- Are Chain-Ladder assumptions violated by the detailed data?

<u>linear regression in R:</u>

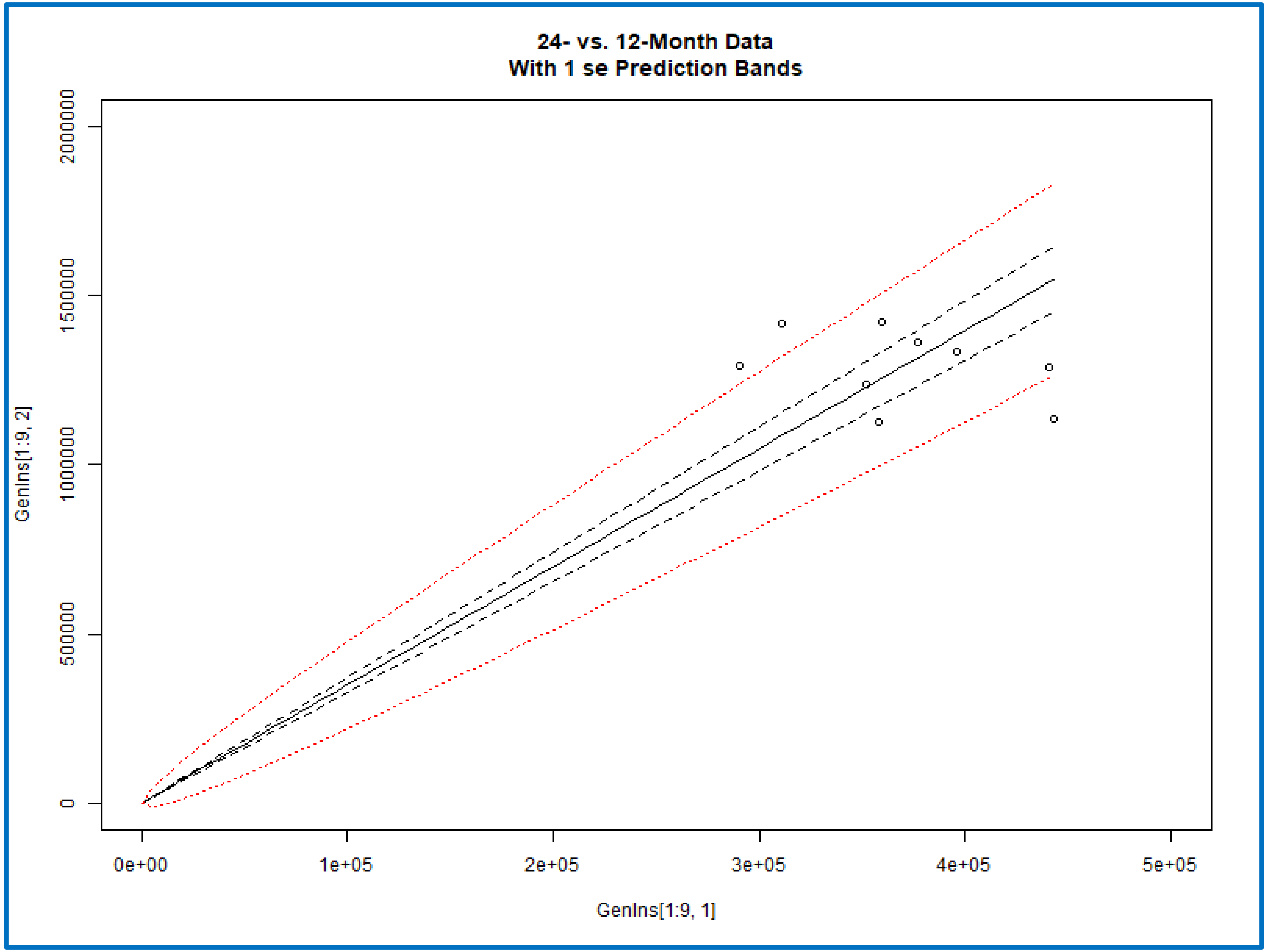
```
lm(y~x)
Coefficients:
(Intercept)    x
1330.6        0.96
```



The 12-24 month relationship from the GenIns triangle

- Are Chain-Ladder assumptions violated by the aggregate data?

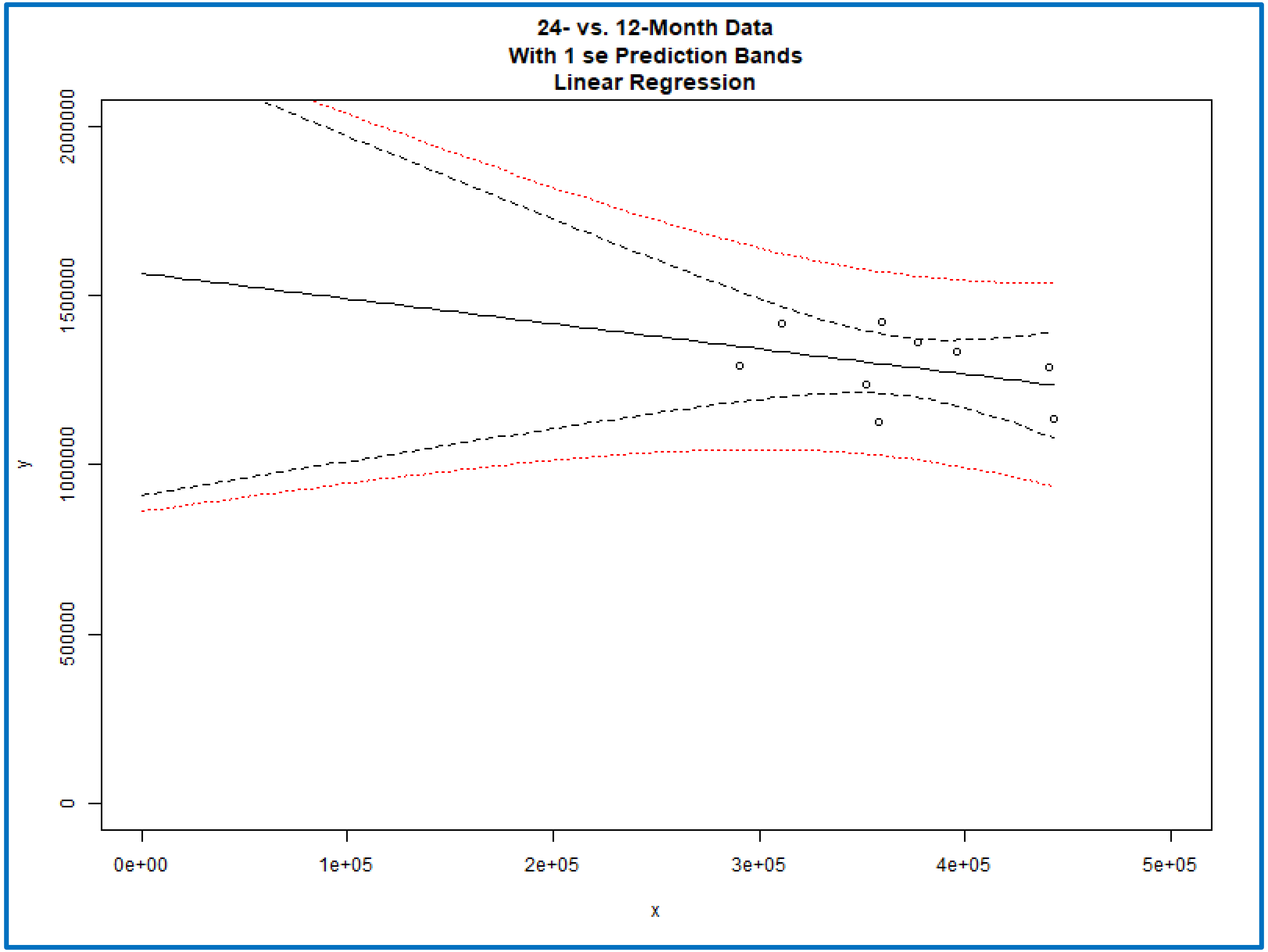
Is that the line you would draw through that data?



The 12-24 month relationship from the GenIns triangle

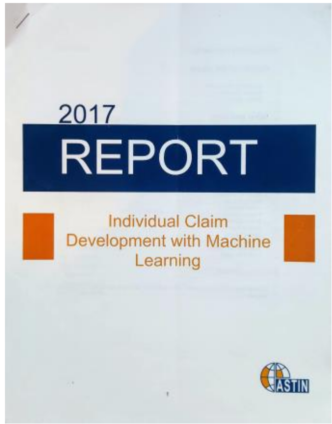
- Rhetorical question:

  Why should this model not be considered for projecting the 12-month value?

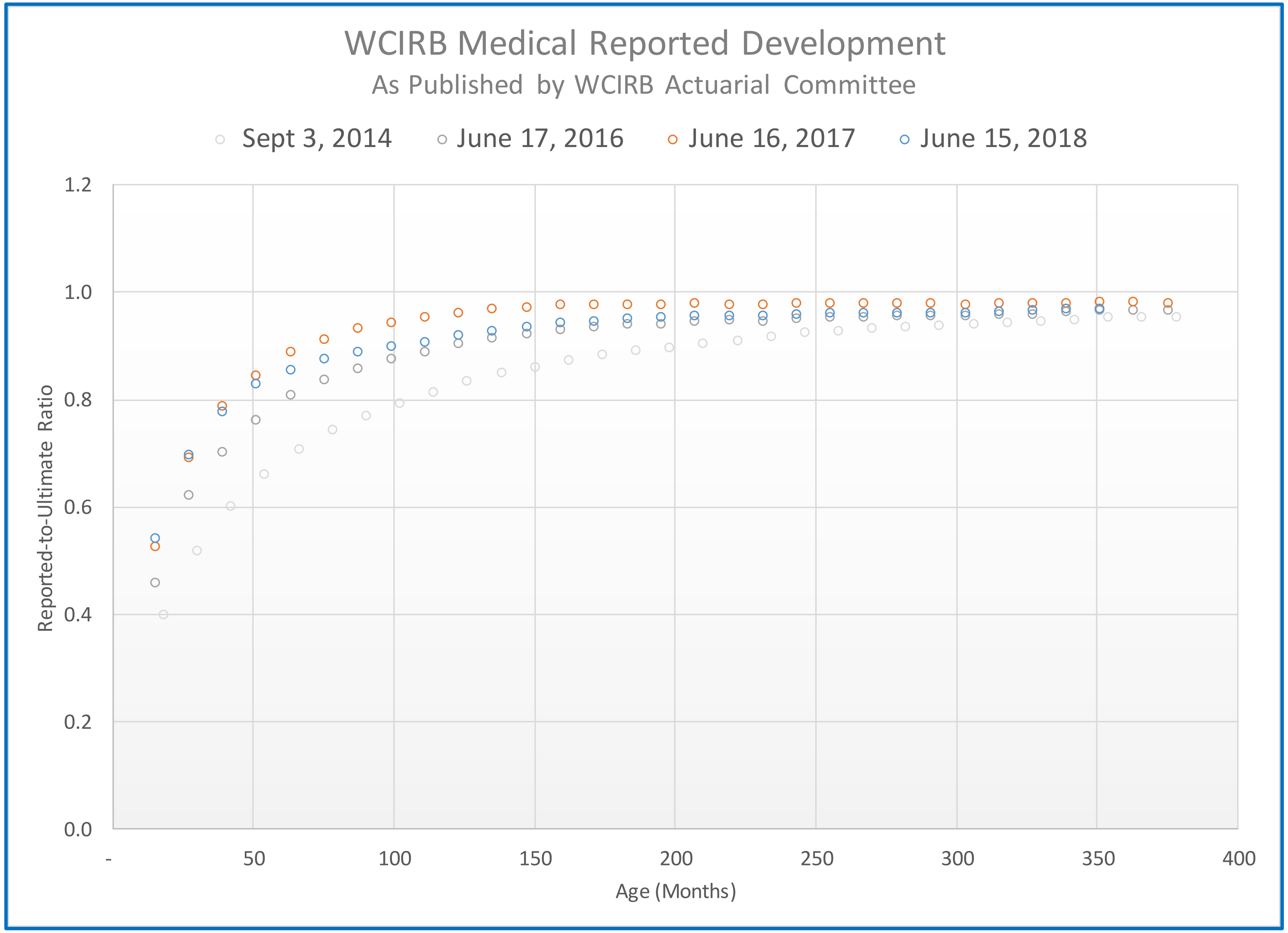


What's next?

- How to model serial correlation?
  - ARMA
  - Michael Wacek, "The Path of the Ultimate Loss Ratio Estimate", *eForum*
- Growth curves
  - Sherman; Clark; Guszcza
- Bayes
- Wüthrich
  - Individual Claim Development with Machine Learning (2017)
  - Neural Networks Applied to Chain-Ladder Reserving (2018)



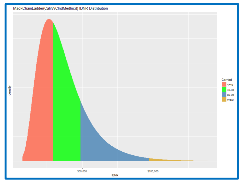
Can InsureTech jump the curve?



* graphics by Kirsten Singer

Summary

- Despite all its problems, the Chain-Ladder Mack/Murphy *model* is *useful*
  - The regression tale of development is easy to understand
  - Distributions help our principals make decisions
- Exciting actuarial analysis in the future
  - Combining methods mid-stream
  - AI modeling of the path to ultimate
- Stories/models with clarity sell best
  - Everybody likes pictures



Q&A

Thank you for coming!

Dan Murphy
dmurphy@trinostics.com