

Surprisals Revisited

Outcome Surprisal

Last year we looked at surprisals for the first time, using them as a way of estimating how unlikely were each weekend's results. You might recall that we measure surprisals in bits and that 1 bit of surprisal is the amount of surprise associated with the win of an even-money favourite, 2 bits of surprisal is the amount of surprise associated with a 3/1 winner, and so on. More generally, n bits of surprisal was the amount of surprise associated with an outcome with a pre-game probability of $1/2^n$.

So, for example, when St Kilda beat Port Adelaide in Round 18 after their respective head-to-head prices were \$1.18 and \$4.50, only 0.35 bits of surprise were generated. St Kilda's implied probability for that game was $4.50/(1.18+4.50) \times 0.99$, or 78.4%, and $-\log_2(0.784)$ is 0.35. (Note that the 0.99 in the probability calculation allows for a 1% probability of a draw.)

Using surprisals in this way was, I think, a useful means of assessing the surprise content of each game's outcome from a head-to-head wagering point of view. Surprise was, it should be noted, quite a black and white affair in that a 1 point win by a very short-priced favourite was considered no more or less surprising than a 40 point win by the same team, but that is a view of surprise consistent with the bookies' treatment of these respective results from a head-to-head wagering point of view.

The black-and-whiteness of the approach to wins and losses aside, draws were problematic for the approach because, at assumed 100/1 odds, they rated as games producing 6.4 bits of surprise. For any team registering a draw, this tended to distort its average surprisal bits per game for the remainder of the season. My usual response to this problem was to leave draws out of any analysis – a time-honoured if not entirely satisfactory behaviour.

To deal with the dilemma of the draw, this year I'm slightly changing the measure, which I'll now call "Outcome Surprisal", by defining the surprisal of a draw as the average of the surprisal associated with the other two results. So, a win by Team A would have produced 0.52 bits of surprisal and a win by Team B would have produced 1.77 bits, then a draw will be defined as producing 1.15 bits of surprisal. I'm no doubt violating some fundamental law of Information Theory by doing this, but pragmatism sometimes deserves a voice.

(I'm reminded of the story about the economist alleged to have proclaimed "Well now we know it works in practice; I wonder if it works in theory too?")

Anyway, let's have a look at last year's results (including the Finals) through the lens of this new Outcome Surprisal measure.

Team	Total Surprisals from ...				Average Surprisals per ...				Rank of Average Surprisals		
	Wins	Losses	Draws	Total	Win	Loss	Draw	Game	Wins	Losses	Total
Adelaide	9.90	9.36	0.00	19.25	0.76	0.94		0.84	12	8	10
Brisbane Lions	6.59	11.89	0.00	18.47	0.66	0.99		0.84	13	6	8
Carlton	13.88	10.23	0.00	24.11	1.39	0.85		1.10	3	12	1
Collingwood	11.73	14.07	0.00	25.79	0.90	1.28		1.07	9	3	2
Essendon	9.68	7.84	0.00	17.52	1.21	0.56		0.80	4	15	12
Fremantle	6.20	15.24	0.00	21.44	1.03	0.95		0.97	7	7	5
Geelong	7.75	3.80	0.00	11.54	0.34	1.90		0.46	16	1	16
Hawthorn	11.44	8.64	0.00	20.08	0.57	1.73		0.80	15	2	11
Kangaroos	13.43	8.66	1.02	23.11	1.12	0.87	1.02	1.00	6	10	4
Melbourne	5.34	6.72	0.00	12.06	1.78	0.35		0.55	1	16	15
Port Adelaide	7.20	14.92	0.00	22.13	1.03	0.99		1.01	8	5	3
Richmond	12.64	6.09	1.14	19.87	1.15	0.61	1.14	0.90	5	13	6
St Kilda	11.52	9.44	0.00	20.96	0.82	0.86		0.84	10	11	9
Sydney	8.13	8.85	1.02	17.99	0.63	0.88	1.02	0.75	14	9	14
West Coast	6.41	10.31	0.00	16.71	1.60	0.57		0.76	2	14	13
Western Bulldogs	12.46	8.25	1.14	21.85	0.78	1.03	1.14	0.87	11	4	7
Total	154.29	154.29	4.32	312.90	0.84	0.84	1.08	0.85			

The last three columns are the ones on which to focus. From them you can conclude that, on average, Melbourne's wins were the most surprising, West Coast's were next-most surprising, Geelong's were least surprising, and Hawthorn's were second-least.

Looking next at losses, Geelong's were the most surprising, Hawthorn's the second-most, Melbourne's were least surprising and Essendon's second-least.

Overall, Carlton's results were the most surprising, Collingwood's second-most, Geelong's least and Melbourne's second-least. Given that analysis, you'd expect BKB to have done poorly in tipping Carlton and Collingwood, and well in tipping Geelong and Melbourne. That was indeed the case, with BKB correctly predicting the results for Collingwood and Carlton in just 13 of the 22 home-and-away season games, and correctly predicting 21 of Geelong's and 18 of Melbourne's 22 results.

So, the Outcome Surprisal measure seems to me to be a reasonably summary of the head-to-head aspects of the season.

Next, let's take a look at 2008 from a round-by-round perspective.

Round	Ave Surprisals	Rank	Designation
1	0.84	11	Predictable
2	0.75	17	Very Predictable
3	0.83	13	Predictable
4	1.10	4	Unpredictable
5	0.78	16	Predictable
6	0.56	24	Very Predictable
7	0.88	10	Somewhat Predictable
8	0.55	25	Very Predictable
9	1.16	3	Unpredictable
10	0.57	23	Very Predictable
11	0.67	21	Very Predictable
12	0.81	14	Predictable
13	1.09	5	Unpredictable
14	1.07	7	Unpredictable
15	0.72	19	Very Predictable
16	0.94	9	Somewhat Predictable
17	1.38	2	Unpredictable
18	0.62	22	Very Predictable
19	0.78	15	Predictable
20	1.05	8	Unpredictable
21	0.67	20	Very Predictable
22	0.84	12	Predictable
23	0.73	18	Very Predictable
24	1.09	6	Unpredictable
25	0.35	26	Very Predictable
26	1.51	1	Unpredictable
Total	0.85		

Given that the only change we've made is to include draws using the new definition described above, it's not surprising that this table (and the earlier one for the teams) looks much like the table we produced last year.

In terms of average surprisals per game, the Grand Final proved to be the most surprising round of the year. Curiously, the previous round had provided the season's least surprising head-to-head results.

Other relatively surprising rounds were Round 17 (which included Essendon's win at \$3.45 against Collingwood and West Coast's win at \$5 against St Kilda), Round 9 (which included Collingwood's win at \$4.50 against Geelong and West Coast's win at \$3.60 against Adelaide) and Round 4 (which included Richmond's win at \$5.50 against Fremantle and Carlton's win at \$4.00 against Collingwood).

Investigating the table a bit more deeply, it turns out that results generally became more surprising as the season progressed in 2008. As you can see from the table at right, this had not been the trend in either of seasons 2006 or 2007.

Rounds	2006	2007	2008
1-11	0.90	0.96	0.79
12-22	0.87	0.89	0.90
Finals	0.86	0.79	0.92
Total	0.89	0.92	0.85

Margin Surprisal

In the previous section we looked at how surprising were head-to-head results, using the head-to-head market prices to infer probabilities and, from these, the surprisal of the result.

There are other aspects of a result for which we might also want to assess surprise levels. One such aspect is the margin of victory, and it's this aspect that I also intend to analyse this year. More precisely, I'll be analysing the absolute margin of victory adjusted for the points start being offered by the favourite – what I'll call the Absolute Handicap-Adjusted Margin (AHAM). To calculate the AHAM for a game we take the Favourite's score subtract the Underdog's score and add the points handicap (taken from the Favourite's perspective, and so a negative number), then take the absolute value of the number we obtain.

So, if Sydney beat Brisbane 90 points to 60 points and Brisbane were 12½ point favourites, the AHAM for the game would be the absolute value of 60-90-12½, which is 42½ points.

You can think of the AHAM as a measure of the accuracy of the bookies' handicapping. Where the AHAM is small, the bookies have done a good job in estimating the relative strengths of the two teams and setting an appropriate handicap; where the AHAM is large they've done a poor job. Last year, the average AHAM was just 29.8 points per game, which is further evidence that the bookies know a thing or two about football.

To calculate the surprisal value of any given AHAM, we need probabilities. While there's no betting market for AHAM we can derive empirical probabilities using historical data, and this is what I've done in the table at right, using information for all games from the previous three seasons.

AHAM	Prob	Surprisals
<i>Less than 12</i>	26.1%	1.94
<i>12 < 24</i>	23.9%	2.07
<i>24 < 36</i>	17.2%	2.54
<i>36 < 48</i>	14.7%	2.77
<i>48 < 60</i>	9.0%	3.48
<i>60 < 72</i>	5.1%	4.28
<i>72 < 84</i>	2.4%	5.39
<i>85 < 96</i>	1.7%	5.92
<i>97 < 108</i>	1.1%	6.51
<i>108 or more</i>	0.7%	7.09

Applying this table to the results of the 2008 season, we get the following.

Team	Total Surprisals from Handicap			Ave Surprisals per Handicap			Rank of Ave Surprisals for Handicap		
	Wins	Losses	Total	Win	Loss	Game	Wins	Losses	Total
Adelaide	27.55	32.84	60.39	2.76	2.53	2.63	7	11	9
Brisbane Lions	18.78	30.35	49.13	2.09	2.33	2.23	16	13	16
Carlton	31.93	25.70	57.63	2.66	2.57	2.62	10	10	10
Collingwood	40.42	30.15	70.57	2.89	3.02	2.94	2	3	2
Essendon	25.49	36.96	62.45	2.83	2.84	2.84	5	5	4
Fremantle	28.17	27.67	55.85	2.35	2.77	2.54	15	8	12
Geelong	45.74	25.12	70.85	2.86	2.79	2.83	3	7	5
Hawthorn	46.69	19.84	66.53	2.59	2.83	2.66	11	6	8
Kangaroos	31.29	28.48	59.77	2.41	2.85	2.60	13	4	11
Melbourne	21.27	41.22	62.48	2.36	3.17	2.84	14	1	3
Port Adelaide	30.06	29.04	59.10	2.73	2.64	2.69	9	9	7
Richmond	34.23	27.51	61.75	3.11	2.50	2.81	1	12	6
St Kilda	30.69	32.06	62.75	2.79	2.29	2.51	6	14	13
Sydney	27.45	31.35	58.80	2.74	2.24	2.45	8	16	14
West Coast	17.01	49.29	66.30	2.83	3.08	3.01	4	2	1
Western Bulldogs	35.97	25.15	61.12	2.57	2.29	2.44	12	15	15

Again, it's the final three columns that tell the story.

Richmond, it turns out, had, on average, the most surprising handicap wins, driven largely by their Round 4 win by 64 points over Fremantle when they were receiving 37½ points start (giving an AHAM for this game of 101½ points and a surprisal score of 6.51), and their 77 point win over West Coast in Round 15 when narrow favourites (producing 5.39 bits of additional surprisal). Next, in terms of surprising handicap wins, came Collingwood and Geelong.

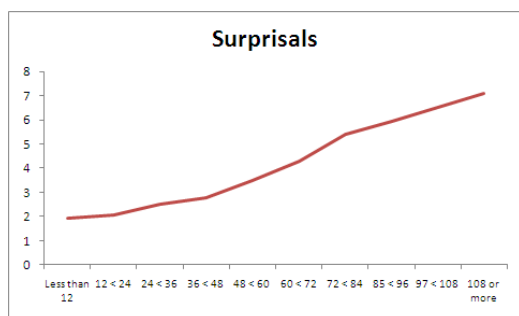
Melbourne had the most surprising average handicap losses, which included a 104 point loss to Hawthorn in Round 1 when receiving 20½ points start (5.39 bits), a 95 point loss to the Bulldogs in Round 2 when receiving 18½ points start (another 5.39 bits), and a 78 point loss to Port in Round 21 when receiving 15½ points start (4.28 bits). Next most surprising were West Coast and Collingwood.

Overall, the most surprising results based on AHAM belonged to West Coast. On average, each game they played generated just over 3 bits of surprisal. The second-most surprising set of handicap results belonged to Collingwood, and third-most came from Melbourne.

You'd be forgiven for thinking that a team's ranking on Margin Surprisals might be equivalent to its ranking on average AHAM per game. And, as the table at right shows, that forgiveness would be well-deserved. What's particularly startling to me about this table is the magnitude of the difference in average AHAM between the highest and lowest ranked teams. Collingwood's 36.71 point average is 1.9 times that of Brisbane Lions.

Team	Ave AHAM	Rank	Rank on Surprisals
Adelaide	27.93	11	9
Brisbane Lions	19.32	16	16
Carlton	30.64	8	10
Collingwood	36.71	1	2
Essendon	35.14	3	4
Fremantle	26.36	13	12
Geelong	32.38	5	5
Hawthorn	30.78	7	8
Kangaroos	27.07	12	11
Melbourne	32.27	6	3
Port Adelaide	30.00	9	7
Richmond	32.45	4	6
St Kilda	28.02	10	13
Sydney	26.17	14	14
West Coast	35.55	2	1
Western Bulldogs	25.74	15	15
Total	30.20		

What differences there are in the average AHAM and Surprisals rankings stem from the fact that the relationship between an AHAM and its Surprisal value is not strictly linear, a fact that you can see for yourself by inspecting the table above linking AHAM ranges to Margin Surprisals, or by sneaking a peak at the chart below.



In particular, large AHAMs, being proportionately rarer, generate higher levels of surprise than a strictly linear relationship would suggest.

This non-linearity tends to lift the Surprisal-based ranking of teams that are associated with large AHAMs such as West Coast and Melbourne (see table below).

So, West Coast grabs 1st place on Margin Surprisals partly because of the number of times it was associated with AHAMs of 72 or more (4 times), and partly because of the rarity with which it was associated with AHAMs below 12 points.

Team	0<12	12<24	24<36	36<48	48<60	60<72	72<84	96<108	108 or more
Adelaide	7	4	1	7	2	1	1	0	0
Brisbane Lions	10	5	3	3	1	0	0	0	0
Carlton	4	5	3	4	6	0	0	0	0
Collingwood	3	6	2	6	4	1	1	0	1
Essendon	5	6	0	2	6	2	1	0	0
Fremantle	5	7	5	2	2	0	0	1	0
Geelong	5	6	6	4	1	0	1	1	1
Hawthorn	6	4	5	5	3	1	1	0	0
Kangaroos	6	6	6	1	2	1	0	1	0
Melbourne	6	1	7	2	3	1	2	0	0
Port Adelaide	5	6	4	3	1	2	0	1	0
Richmond	5	5	3	3	4	0	1	1	0
St Kilda	6	7	4	3	4	1	0	0	0
Sydney	6	7	3	5	2	1	0	0	0
West Coast	2	8	5	1	1	1	3	1	0
Western Bulldogs	9	5	5	3	2	0	1	0	0

Were you a margin bettor last year, the Lions would have been the scariest team to back or lay. On 10 occasions their margin bets were decided by less than 2 goals. The Bulldogs weren't far behind them, producing 2 goal or less margins on 9 occasions.

Finally, let's have a look at the round-by-round view of Margin Surprisals (see table at right).

Round 2, it turns out, produced the season's most surprising set of AHAMs. This round included four games with AHAMs of 60 points or more. Round 22 produced the second-most surprising set of AHAMs and included four games with AHAMs of 48 points or more (one of which was 105½ points).

The season's least surprising set of AHAMs came in Round 5, for which the highest AHAM was just 23½ points. Next least surprising was Round 18, which included just 2 games with an AHAM greater than 21½ points.

Across the entire season, the average Margin Surprisals per game in 2008 was 2.66. This is slightly higher than the figure of 2.63 for season 2007, and slightly lower than the figure of 2.73 for season 2006.

Round	Ave Surps	Rank
1	2.70	11
2	3.49	1
3	2.28	24
4	3.06	5
5	2.04	26
6	2.31	21
7	2.47	17
8	2.33	20
9	3.15	3
10	3.05	6
11	2.36	19
12	2.52	16
13	3.07	4
14	2.45	18
15	2.74	9
16	2.31	22
17	2.82	7
18	2.17	25
19	2.64	13
20	2.73	10
21	2.55	15
22	3.47	2
23	2.59	14
24	2.65	12
25	2.30	23
26	2.77	8
Total	2.66	

Summary

So, this year we'll be tracking the level of surprise associated with two aspects of each game: the head-to-head result relative to the bookies' odds, and the absolute handicap-adjusted margin of victory relative to probabilities based on empirical data.

Last year, the top and bottom rankings for Most Surprising Head-to-Head Results and Most Surprising Absolute Handicap-Adjusted Margin Results were as follows:

Ranking	Head-to-Head Surprisals	AHAM Surprisals
1 st	Carlton	West Coast
2 nd	Collingwood	Collingwood
3 rd	Port Adelaide	Melbourne
4 th	Kangaroos	Essendon
5 th	Fremantle	Geelong
Spoon	Geelong	Brisbane Lions

Geelong's head-to-head results were not what was surprising about the Cats last year; it was, to a larger extent, the size of their victories (and rare losses) relative to what the bookies were expecting.

Collingwood, on the other hand, were surprising no matter how you looked at them. They often won when they were expected to lose, and lost when they were expected to win. What's more, the size of those wins and losses surprised even the bookies.

24 January 2009