

History shows the polls don't lie (II)

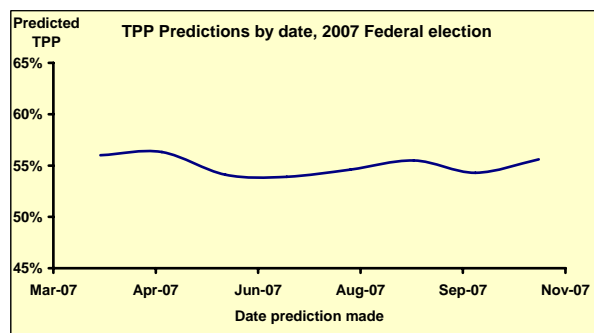
Summary

History shows the polls don't lie. That was a title of an article I wrote for the Australian Financial Review in the lead-up to the 1996 election, in which I predicted a 50-seat majority for John Howard. After this duly occurred, the AFR asked me to explain how I did it and this duly appeared as *Maligned opinion polls got it right*. I think that both these articles can be found in the archives of Peter Brent's Mumble website. Peter himself has claimed prediction bragging rights arising from his own accurate forecast for the 2004 election and has issued a challenge for us all to put up or shut-up, making a prediction of 90 seats for the ALP, one more than the "bold" Malcolm Mackerras. Poll Bludger has gone for 84. Below I suggest 97 is the "most likely", though random variation could produce anything from. 87 to 104.

"Everyone knows" that polls are "not predictive" and, as Mumble points out, Gerard Henderson will rub your face in the egg if you come a gutser trying to prove otherwise. The perception that the polls lie arises from a selective reading of immediate pre-election polls at past elections. Although the vagaries of sampling error are routinely acknowledged, they are often forgotten when pundits look at the polls which emerge on the Thursday before the election. In this way, reputations are made or broken on the strength of random sampling error.

As *History shows the polls don't lie* explained, the polls need to be "coddled". In this case, coddling means aggregating several successive polls from as many pollsters as possible, segregating the results in an appropriate way, analysing the results over long-enough time periods to detect any underlying time-dependent changes and applying the results on an electorate-by-electorate basis. There is plenty of data going back to the 1946 election which allows one to apply these techniques retrospectively and reach an assessment of whether they work. They usually do.

Applying this sort of poll analysis to the polls from mid 2006 to early November 2007 strongly suggests a Labor victory, with an Australia-wide two-party preferred vote for the ALP of 54.7%-56.3%, most likely 55.5% and a majority of 24 to 58 seats, most likely about 44 seats. The methods elaborated below have pointed consistently at a result close to this since April 2007.



TPP

The Two Candidate Preferred (TCP) vote is the sole determinant in an election for a seat. Expressed as a percentage it is the proportion of all ballot papers upon which the number allocated to one TCP candidate by the voters is a lower number than that allocated to the other TCP candidate. The TCP is not formally counted this way, but the foregoing method is functionally identical to the process actually used. The TCP votes of the two candidates necessarily add to 100%. Because two parties dominate politics, the TCP vote is nearly always a *TPP (Two Party Preferred)* vote and, by long-held convention, is always expressed as an ALP TPP. In those cases where a

contest is TCP rather than TPP, the AEC will usually provide an ALP TPP subsequent to the election, for information purposes.

In opinion polls and at elections, (i) an Australia-wide TPP; (ii) a State-wide TPP or; (iii) a TPP for a particular class of electorate may be calculated by a number of different averaging techniques. The focus of most psephologists is the Australia-wide TPP. Usually, but not always, an Australia-wide TPP of greater than 50% means that a party will obtain a majority of seats and so form a Government. Concentration of extreme TPPs into certain electorates, gerrymanders and the fact that some electorates vary unavoidably in size from the average, means that parties can sometimes win Government, or fail to win Government, when their nation-wide TPP indicates otherwise. It would be possible in an extreme case for a party with a little more than 25% of the TPP to form a majority Government. Historically, the nett effect of these influences has somewhat disadvantaged the ALP, which needs nearly 51% of the nation-wide TPP to win 50% of the seats. This number is derived from the analysis of elections since the Second World War and is illustrated by the left-hand chart under *Pendulum Performance*, below

Swings

Without *pendulums*, there would be no such a thing as a *swing*. The pendulum is nearly always calibrated in the amount by which the TPP in an electorate deviates from 50%. At an election, the “swing” is the change in TPP% between one election (possibly redistributed) and the next. Swings are usually more convenient than the resultant TPP in visualising the movement of voter sentiment.

Predictive accuracy of the polls

In the 24 elections from 1946-2004, projections made from the aggregated polls produced an average overestimate in predictions for the ALP TPP of 0.6% (graph below). In the worst case (1987), the error in the TPP was 3.8%. In half the elections, the error was less than 1%. There are differences among pollsters (the “house effect”), but these tend to cancel one another out, so that the average of all pollsters has nearly always proved to be the best estimator.

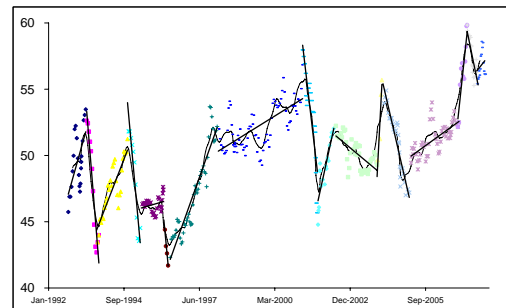
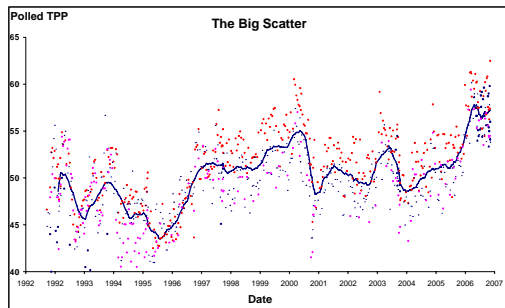
Poll trends

The TPP vote revealed by polling is rarely static. Variation occurs in the apparent numbers from week-to-week and between pollsters. The greater part of this variability is due to the fact that only a small proportion of the voters are ever sampled (usually 1,000-2,000). The “luck of the draw” means that the TPP derived from some samples will deviate from the true nation-wide TPP. The sampling is, strictly speaking, *hypergeometric*, but is usually assumed to be *binomial*, so binomial probability theory can be used to calculate a *margin of error*, which usually means the *95% confidence limits*. The margin of error can be reduced by polling more people or by lumping together the polls from several pollsters and/or several polling periods. Either *raw* or *lumped* numbers are often plotted on graphs, and often with *smoothed* lines connecting the dots, to reveal the underlying *trends*. This is what the *Reuters Poll Trend* does- they have been using the method for a decade and I rather like to think they got the idea from me. I have been using it in a non-predictive way since 1987. Reuters use a particular technique, developed for actuarial purposes better suited to periodically repeating data. Simple moving averages are sufficient. But trend lines without all the data points (*scatter-plots*) from which they are derived tend to induce

the viewer to perceive patterns which the statistics show do not exist. It is possible to see patterns in scatter plots but only in the longer term- say a year or more.

Trends persist for some time, on average about 6 months, before a *tip-point* occurs. Tip-points can usually be ascribed to (or at least triggered by) a political event such as the 1991 recession, the 2001 petrol-price back-flip and the ascendancy of Kevin Rudd. Between tip-points, statistical analyses show that trends are best modelled by assuming that the drift in TPP is steady (linear) with time. The period of longest persistence of drift in recent times was the four years from 1997 to mid-2001. In all except the 2001 election, it has not been necessary to model the trend as anything other than a single straight line. The 2001 election can be modelled with two successive straight lines pre- and post-Tampa, each of which predicted the result to better than 1% accuracy.

The two graphs below show scatter-plots for 15 years of polling data. One has been fitted with a six-month moving average trend line, the other has in addition, a series of straight lines which fit the data.



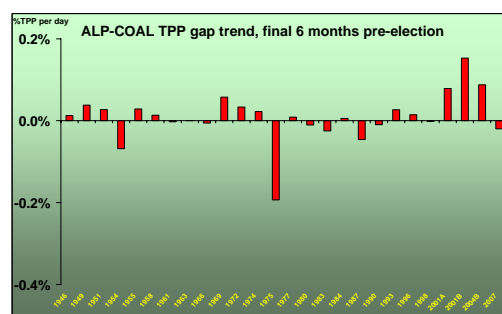
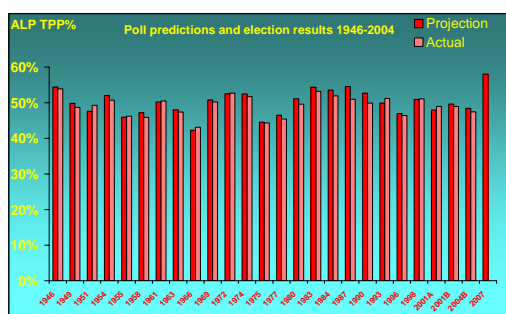
The Gap always narrows.

The difference between the TPPs of the two major party groupings is usually referred to as the *gap* and sometimes as the *lead*. The “gap always narrows during an election campaign” is not strictly true. In about 60% of elections it has, but in 40% it has widened. On average, the ALP has been the beneficiary of gap closing, but the nett benefit to it has been very small. Any change in the gap during an election campaign is usually similar to that which had previously occurred in the 6-month pre-election period. In only 5 of the last 24 elections did the gap change from narrowing to widening or *vice versa*. On average, the gap has widened by about 0.06% per week in the pre-campaign period, but narrowed by 0.07% per week in the campaign period. These means are not statistically different from each other nor from zero.

	1946-2004 mean measures of change in absolute ALP TPP or TPP gap					
	(where ALP declines)		(where gap narrows)		Max TPP slope	Max. gap slope
	Averages		Gap change			
Long-term slope	0.0017%	-0.056%	0.0093%	-0.040%	-0.18%	-0.16%
Campaign-only slope	-0.0010%	-0.052%	-0.0064%	-0.068%	-0.16%	-0.21%
Average error in projection	0.64%		0.57%		(Slopes in % TPP per day)	
Std Dev	1.18%		2.84%			

It is possible to model any election as though it were an average with the above characteristics. During 2007 it has seemed that the election has been significantly different from the long-term average, with the slope being near the top of the range observed since 1946.

Polling accuracy and the behaviour of the gap 1946-2007 are shown below



Voter behaviour

It really does seem that there is such a thing as a *rusted-on* voter and that this is probably a genetic trait. It seems that about 32% of voters are rusted-on to the each of the Coalition and the ALP. These numbers are derived from polling and election statistics by consideration of the extreme ranges and standard deviations seen in either the polling or the voting figures. Some people, even John Howard, think this number is slowly declining with time. Voters can become rusted-on to other parties, but most of these parties have too short a life to be able to say for certain. The Democrats, for instance, seem to have a 1% rusted-on vote. The Greens might have a higher level- but probably not. In all, no more than 5% of voters could conceivably be rusted on to *Others*.

It seems that there is also probably such a thing as a *rusted-on preference voter*- the voter who always preferences the two major party groupings in a certain order, no matter how far down the preference list they put each of them. The value of this statistic appears to be about 40% for both the major party groups.

Taking away the rusted-ons leaves some 30% of voters whose primaries swing and 20% of voters whose preferences swing. It is obviously these voters who are the natural targets in an election campaign. While the national mood probably affects all voters, it is only these 20 or 30% who are likely to change their primary or preference voter as a result of its influence.

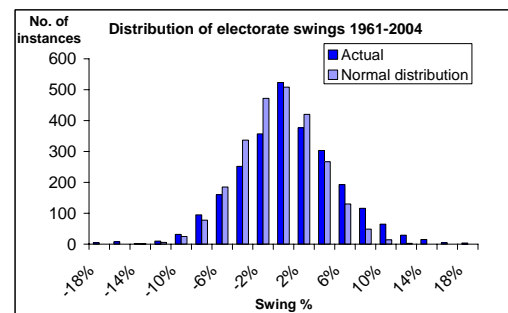
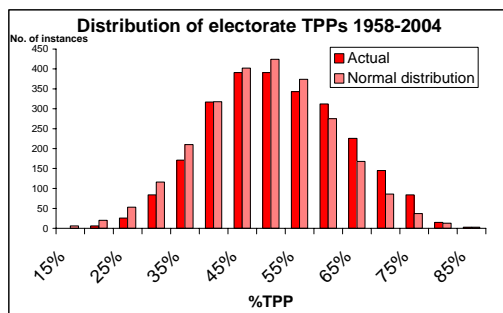
Electorate behaviour

It is possible to analyse past elections to make estimates of how electorates “behave” at elections. It is natural to assume that this behaviour could be better understood if one had precise knowledge of the political pressures at work in each electorate. This is certainly true of the national vote for some indicators (such as interest rates). While this is in theory probably true for electorates too, in practice these pressures are diverse, contradictory, hard to discern and harder to measure. Particularly in marginal electorates, the campaign efforts by major parties seem to cancel one another out to a variable extent. It therefore usually suffices to treat the numbers as though they were samples drawn from populations having certain characteristics which vary according to some statistical distribution. I call this method *phenomenological psephology*; it is just tea-leaf reading or, as someone called it “silly mathematical games”, but it has the virtue of working. The statistics of most interest for both the TPP and the swing are the *mean* and the *standard deviation*.

From analysing the TPP at all elections and for all electorates, it can be discerned that electorate-wide TPPs are distributed in a manner that does not differ substantially

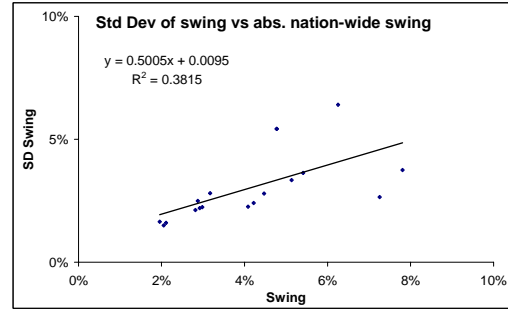
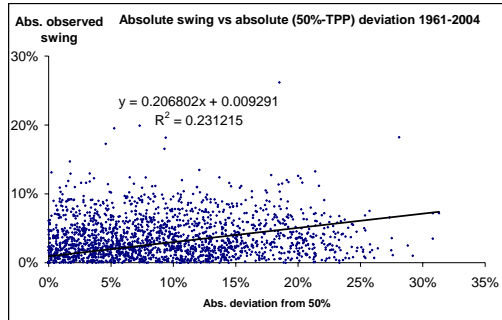
from a *normal distribution*—except in those cases where the TPP varies more than 25% from the mean. The number of instances of this is very small- a handful out of the 2,500 which are in the data-set for 1958-2004. The mean ALP TPP is a little below 50% for this data set, indicating merely that the Coalition has won more elections than has the ALP in this time. The standard deviation of the TPP for this data set is about 12%, meaning that 95% of all measured TPPs ought to fall between 26% and 73% ALP TPP. In fact, 96% fall between these limits, a satisfying corroboration of *normality*.

Analysis of electorate swings produces similar statistics- a near normal distribution, but slightly both *platykurtic* and *leptokurtic*, indicating that swings tend to cluster rather too closely about the mean (0.4%) and in the extremes, leaving bare shoulders. The standard deviation of the absolute swing is 3.6% overall but when analysed on a single-election basis can be shown to be dependant upon the Australia-wide swing, rising with the latter from about 2.5% where there is no nation-wide swing, to 5.9% where the swing is 7.5%, the highest ever seen in the 1961-2004 period.



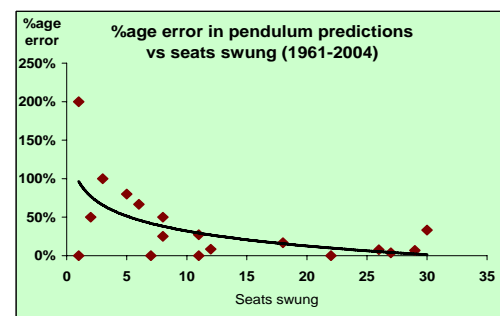
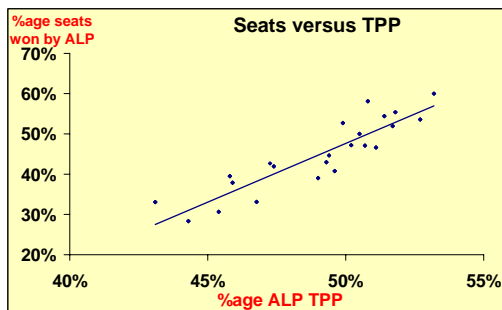
The effect of *marginality* on electorate swings

About 34% of Federal seats have been *marginal* in the 1958-2004 period. The AEC defines a marginal seat as one with an ALP TPP of 44% to 56%. Not everybody may agree on these numbers, but they usually agree that “elections are won and lost in the marginals”. In theory this again is true but, in practice the average swing in the marginal seats rarely differs from the nation-wide swing. The scatter-plot below shows how the absolute swing has varied with the marginality in over 2,400 electorate results (where *marginality* = deviation from 50% TPP). The correlation is positive because of a few data points with high swings, but the slope is small. At any rate, the idea that bigger swings occur in marginal electorates is not supported by the data. If any idea is supported it is that swings are bigger in the safer seats. Thus, if campaign effort has been concentrated in the marginal seats over the years, then the efforts of both sides must have cancelled at almost every election. Only 1998 bucks this trend. Electorate swings are also *homoscedastic* when plotted against TPPs- that is; marginal electorates are no more volatile than are non-marginal electorates. The right hand graph shows how the variability (standard deviation) of electorate swings is dependent upon the size of the nation-wide swing- that is elections are *heteroscedastic* on an Australia-wide basis.



Pendulum performance

The pendulum, that popular weapon of the psephologist is a very useful predictive device at election time. The idea is that one ranks electorates along its path in TPP order and counts how many electorates will come within the range of a particular Australia-wide swing. It is admitted by its proponents that swings vary from electorate to electorate but the proponents argue that any errors arising from this tend to cancel out, so that the overall prediction is likely to be nearly right. Unfortunately for this idea, the pendulum has proven to be a rather poor predictor of nett seat change when the potential seat change is fewer than 10 seats. The performance in both a relative and an absolute sense is pretty good above a 10 seat change. Given the usual statistics on swings, one can conduct a *Monte Carlo simulation* to model the pendulum's performance. For the current election with a swing of (say) 8%, where the expected seat change is about 37 towards the ALP, repeated simulations show that the nett seat change should range from 30 to 45 in 19 out of 20 such elections. Up to 2 ALP seats could be lost under these same statistical conditions. This range is purely due to random statistical fluctuation and does not indicate a pendulum failure.



Using the polls to predict an election result

The technique I have used with success in the past is to use a number of different analyses of polling results to arrive at an anticipated swing for each electorate. For some electorates, the only swing available is that arising from the nation-wide polling estimates. Although I have shown that swings in an election are distributed normally, there is still insight to be gleaned from the polling estimates potential size of swing in each seat individually. This may come from polling in individual seats or from disaggregation of nation-wide polls into different classes of seats, such as by State or marginality. These disaggregations can sometimes reveal potential seat changes that might not be predictable were only an Australia-wide swing to be applied. For most of these statistics one can choose any or all of: (i) the latest number; (ii) the recent medium term average or; (iii) the projection forward to election day. These various methods yield as many as 10 estimates of the swing for each electorate. Not all of these are equally reliable, because of the underlying sample size, because the trends

which lie behind them are non-linear, because of *house effects* of each pollster, or because of other factors. Each number is therefore *weighted* according to its perceived reliability. To some extent the optimal weights for each estimate of swing can be calculated from within the statistics (usually such weights are based on the *variance*), but a leavening of instinct about reliability can also be applied. Assigning a weight of zero to a statistics is equivalent to ignoring it.

Prediction for 24-Nov-2007

Of the methods outlined above, the following have been rejected from the analysis.

The 18-month trend- because it is not linear over that time.

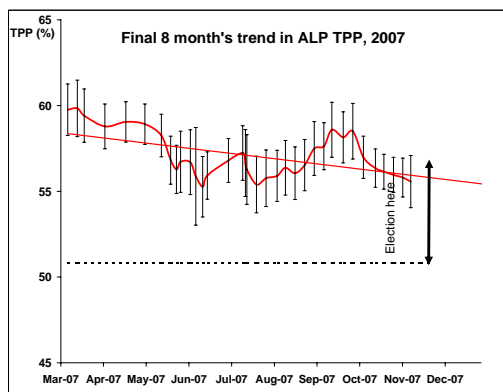
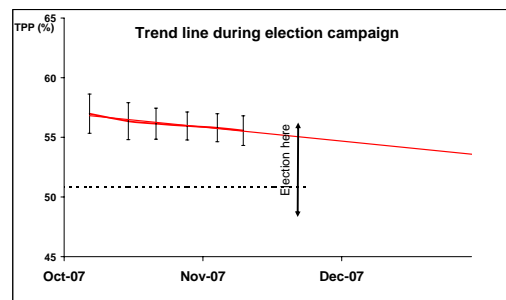
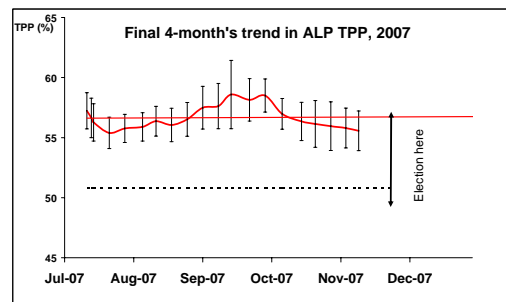
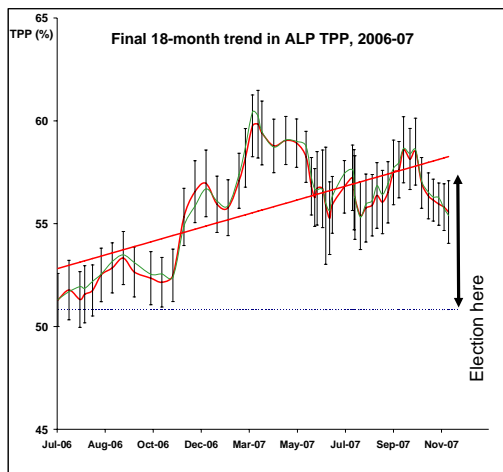
The projection based on this election being “average”.

State-based *projections*- because several states show fluctuations at certain times. I have used only the recent average.

Marginality-based *projections* because complete data is lacking for the marginals post August. Again, I have used only the recent average.

All available polling data has been used, up to the Newspoll poll taken a fortnight before the election.

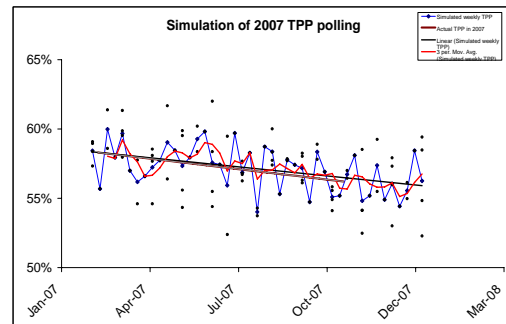
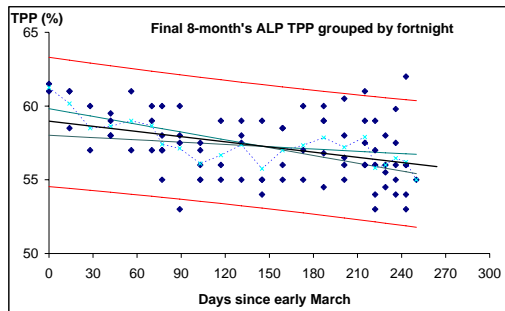
The trend lines look like this:



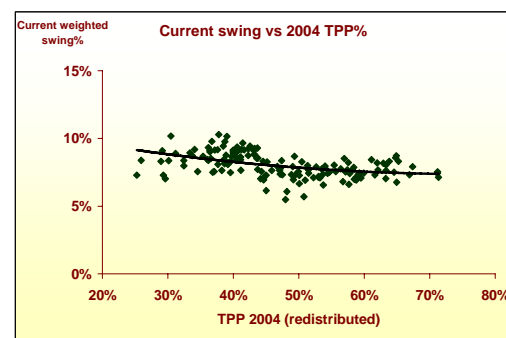
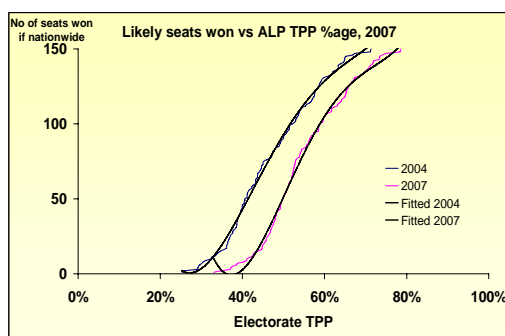
Calculated TPP vote, with sampling error margins, smoothed trend lines and fitted linear models for 4 successively shorter periods preceding the 2007 election. All data except the 18-month projection fit linear models (which should thus remain within the error bars for 95% of the points).

A scatter-plot of the 8-month trend appears below left. The straight line of best fit appears in black and a variance ratio test shows that it is linear with a probability of

greater than 80%. The wasp-waisted lines represent the 95% *fiducial limits* of where the true Australia-wide TPP is likely to lie at any one time point. The more widely-spaced red lines represent the 95% *tolerance limits* for samples making up the line. No more than 5% of the scatter-plot points should fall outside these limits if the data follows a *linear trend*; is *normally distributed* and; is *homoscedastic*. To the right is a typical single Monte Carlo simulation based on polling from 4 pollsters, in which the underlying trend is the same as that in 2007, but every data point has been determined randomly on the basis of a normal distribution. The 2007 linear model and the trend lines (blue-raw; red-smoothed) have been added to show that wobbles like those in the 4 graphs above can be fully mimicked by random fluctuations.



The 8-month, 4-month and campaign projections are all linear and mostly point to nearly the same end result. Each trend line has been given weights according to the closeness of their starting dates to the election. The most recent Australia-wide TPP is used, but given a weighting of only 2. Newspoll polling hints that in 2007, the swing to the ALP is larger in the Coalition non-marginal seats, averaging some 9% (and increasing with time), compared with about 6% in the marginals and 7% in the ALP non-marginals. The projected swing peaks at around 10% in Coalition seats which have a 10% safety margin, making this sort of seat vulnerable. If one gives a high weighting (say, 16) to this observation, up to 9 non-marginal Coalition seats would be lost to the ALP. In the analysis, seat-polling has been given a weight of 2. The graph below, which shows the pendulum in its cumulative form for the 2004 election results and the projected 2007 election results shows that the bigger swing in the safer Coalition seats is hard to discern in the data. It can be seen a little more clearly in the right hand graph, which plots current projected swings against marginality.



Individual seat polling has been given a weight of 4 because there are now a large number of seat polls available- some seats have been polled multiple times and the results are consistent within and between seats. Where no seat polling is available, the

Australia-wide swing is allocated to that seat. Finally, state-based analyses have been given a weighting of 1.

The bottom line (in the table below) is that the weighted national swing is likely to be about 8.2% and the final national TPP about 55.5%- this would give a most likely number of seats for the ALP of 97. The different methods contributing to the weighted results give projected TPPs of 53.8% to 57.3%, and seat numbers ranging from 94 to 102, which is not symmetric about 97 seats, mainly because the clustering of seats on this part of the pendulum is not homogenous. In this region of the pendulum, every 1% swing can produce a 14-seat majority change.

Monte Carlo simulations based on the weighted data show that about 19 out of 20 elections conducted under these conditions would produce a TPP of between 54.7% and 56.3% (55.5% is in the middle of this range) and a 95% confidence limit for the number of ALP seats of 87-104. This is not symmetric, and for the reason cited above..

Method of calculation	Swing	TPP	ALP seats	non-ALP seats	ALP Majority	ALP Seat Gain	Weighting factor
National ALP- recent	8.6%	55.9%	96	54	42	36	2
National ALP TPP Election Projection final 18 months							
National ALP TPP Election Projection final 8 months	8.6%	55.8%	96	54	42	36	2
National ALP TPP Election Projection final 4 months	9.5%	56.7%	102	48	54	42	1
National ALP TPP Election Projection campaign period only	7.8%	55.0%	94	56	38	34	4
National ALP TPP Projection based on historical campaign trends							
National ALP TPP- based on State analyses averages	10.0%	57.3%	102	48	54	42	1
National ALP TPP- based on State analyses Projection							
National ALP TPP- based on Marginality analyses averages	8.1%	55.4%	96	54	42	36	2
National ALP TPP- based on Marginality analyses Projection							
Average of specific electorate polling	6.5%	53.8%	95	55	40	35	2
Weighted average of all methods	8.2%	55.5%	97	53	44	37	

Any projection based on data which is continuous right up until election day cannot *help* but point at the true result. A good test of a predictive theory is whether it can make the same predictions further back in time. The methods described above produced the following estimates throughout 2007.

Date	Swing	TPP	Seats
1-Apr	8.7%	56.0%	100
1-May	8.9%	56.3%	101
1-Jun	6.7%	54.1%	93
1-Jul	6.5%	53.9%	92
1-Aug	7.3%	54.6%	95
1-Sep	7.9%	55.5%	95
1-Oct	6.7%	54.3%	91
1-Nov	8.2%	55.6%	96

It would take a political sensation of biblical proportions for the Coalition to win from here or, as Antony Green has said, *the greatest come-from-behind victory in history*. The chance of the Coalition pulling the ALP TPP vote back to Tiger Territory in the region of 50.6% in the absence of such a sensation, or in the absence of the statistics going pear-shaped, is ridiculously small.