Pottinger
PERSPECTIVES

X1000RPM

START
STOP

ELECTION 2013
PUNDITS V PREDICTIVE STATISTICS

If you believe the current vibe in the newspapers, then a Coalition win at the 2013 Federal Election is a foregone conclusion and the ALP faces electoral annihilation. With inspiration from the success of electoral statisticians such as Nate Silver in the US, we decided to investigate the upcoming election using the statistical armament at our disposal. The conclusion: although the ALP is undoubtedly in trouble, there is still the possibility of a significant ALP recovery (enough to win the election), and the ALP vote is unlikely to get worse from where it currently stands. A modest degree of ALP recovery before the election is the most likely outcome.

The use of statistical methods to aggregate data and make specific predictions about elections has been around for decades. However, the use of statistical techniques has come to public prominence over the last five or so years as a result of the success that these techniques have had in predicting the outcome of US elections.

Although a number of people made predictions about the 2012 US Presidential Election, the person who achieved the most fame was Nate Silver, a statistician who had previously made a name for himself as a sabermetrician (someone who applies statistics to the analysis of baseball). In the 2012 election, his statistical model correctly predicted the outcome of all 50 states in respect of the presidential component of the election, and correctly predicted the outcome in 31 of the 33 Senate elections. Two other academics who had been running well-advertised statistical analyses also predicted the correct outcome (Drew Linzer and Sam Wang), with one also getting the exact outcome.

These techniques were derided by some election pundits, particularly those who were Republican-leaning, who felt that statistical models were too simple to be able to predict the outcome, and that human experience was a necessary intermediary between the data and predictions. There was a significant degree of upset on Election Day when statistical predictions made months in advance performed much better than the pundits did even on the day of the vote itself.

To us, it is no surprise that the statistical models outperformed human judgement. Appropriately constructed models are able to combine large amounts of data in a robust, logical and unbiased framework. The US electoral environment is one such "big data" domain, with many thousands of polls conducted in the lead-up to a US election.

We have created a wide variety of statistical and quantitative models for clients in a number of sectors, including the energy, resources and agricultural sectors. With the 2013 Federal Election in Australia becoming increasingly topical, we couldn't resist applying our statistical toolbox to investigate what might happen over the coming months.

Polling in Australia

Polls are the most visible measure of sentiment throughout the electoral term. They have the advantage of being easy to understand and are published frequently (often fortnightly). Polling results occupy a special place in the minds of local

Appropriately constructed models are able to combine large amounts of data in a robust, logical and unbiased framework

commentators: after politically significant events, the next set of polling results is keenly considered by commentators. These results, and subsequent polls, can be more than enough to sustain the political commentary for weeks.

The quantity of polling in Australia in the lead-up to the election is significantly less than in the US. Critically, in the US there are many state-based polls, which naturally link directly to the state-based electoral college system which determines the President. Polling in Australia, on the other hand, is generally conducted nationally. The outcome

of the Australian House of Representatives is determined on a seat-by-seat basis, and polling information in individual electorates is not generally available, with the exception of a relatively small number of polls conducted in more marginal electorates. The various political parties conduct their own polls in marginal electorates from time to time, but this data is generally closely guarded.

Regular polling in Australia is conducted by a number of organisations, including Newspoll (associated with The Australian), Nielsen (associated with Fairfax publications) and Roy Morgan. A number of other organisations currently conduct federal election polling in Australia, including Galaxy and Essential.

Our model

Our model aggregates available federal polling data to construct an estimate of the two-party preferred (2PP) vote share for the ALP and Coalition on every day between 1 August 2010 and the presumed election day, 14 September 2013.

We assume that voting intention follows a random walk process. This means that the predicted voting intention for the future (given only the current and historic polling data) is on average the same as the estimated value today. The use of this type of model is appropriate if you don't believe in electoral momentum (ie the fact that more people vote for one party today than yesterday means the same is likely to be true tomorrow). This type of model has been used with great success in the US elections.

Given this model for how voter intention changes through time, we calibrate our statistical model against more than 250 polls since 1 August 2010. We use the methods of Bayesian statistics to estimate the parameters of our model given the data

available.

There are two key reasons why the aggregation of polling data in this manner works. The first relates to the idea of sampling error. Because each opinion poll samples only a small fraction of the total population (typically about 1,000 people), the results will differ slightly from the population-wide result. Polling organisations quantify this effect through statements about the "margin of error". This is defined so that if you conduct many opinion polls, the result you would get from polling the entire population should fall within the margin of error of each poll 95% of the time. By aggregating data from multiple polls, you can obtain a more precise estimate of the 2PP vote share than by using the data from just one poll.

The second reason is more subtle – it relates to the fact the people's preferences do not change too rapidly with time. This means that polling information from a week or two ago still has some relevance to estimating the 2PP vote share today. From our model, the estimated average absolute value of the day-to-day change is about 0.2%. Note that this number scales with the square root of time, so the average absolute value of the change over a week is about 0.4%.

Our model estimates the 2PP vote share based on how much information is available at any point. Where more polling data is available, our model naturally produces more precise estimates of what the true 2PP voting intention share is. Where there is less information then our model is more imprecise.

Importantly (and like any good model should), our model makes statements about how precise its predictions are. That is, our model tells us not just that the estimated ALP 2PP vote share is (for instance) 49% on some day, but that there is a 95% chance that the 2PP vote share falls within the range 47.5% to 50.5%. It is just as important to understand how precise your model's predictions are as it is to understand the central estimates.

Voting intention today vs voting intention in September

We have constructed our model using Bayesian statistics for three reasons. The first is that, because we have a model for how voter intentions change through time, we can run our model forward to make predictions about the future. The second is that because we have used Bayesian statistics to create the model, the model makes statements not only around the quantities of interest (eg 2PP vote share) but how precise these estimations are. The third is that Bayesian statistics allows you to combine statistical estimations from one data source together with estimations from another, independent data source to make more precise predictions.

Whilst current voting intentions are interesting, the real question that everyone wants answered is: what will the outcome be on election day? The factors listed above mean that we are in a position to tackle this much more interesting problem directly.

Given the random walk nature of the model, the central prediction on election day will just be the same as the prediction today, which isn't very interesting (or believable). However, things get much more interesting when we start to include other data.

There are a number of other commentators who are using various smoothing techniques to achieve a more precise estimate of today's voting intentions (eg Poll Bludger at Crikey and Pollytics, also at Crikey). Our model for today's voting intentions will not perform significantly better than other smoothing techniques. However, it is difficult to use these local smoothing techniques to make predictions about the future. To the best of our knowledge, no-one in Australia is publicly reporting the results from statistical models created in this fashion which predict the outcome on election day.

Previous elections are a guide to future elections

We can look to historic results as a guide to the likely outcome of this election. We know that elections almost always fall within a fairly narrow range, with the 2PP vote
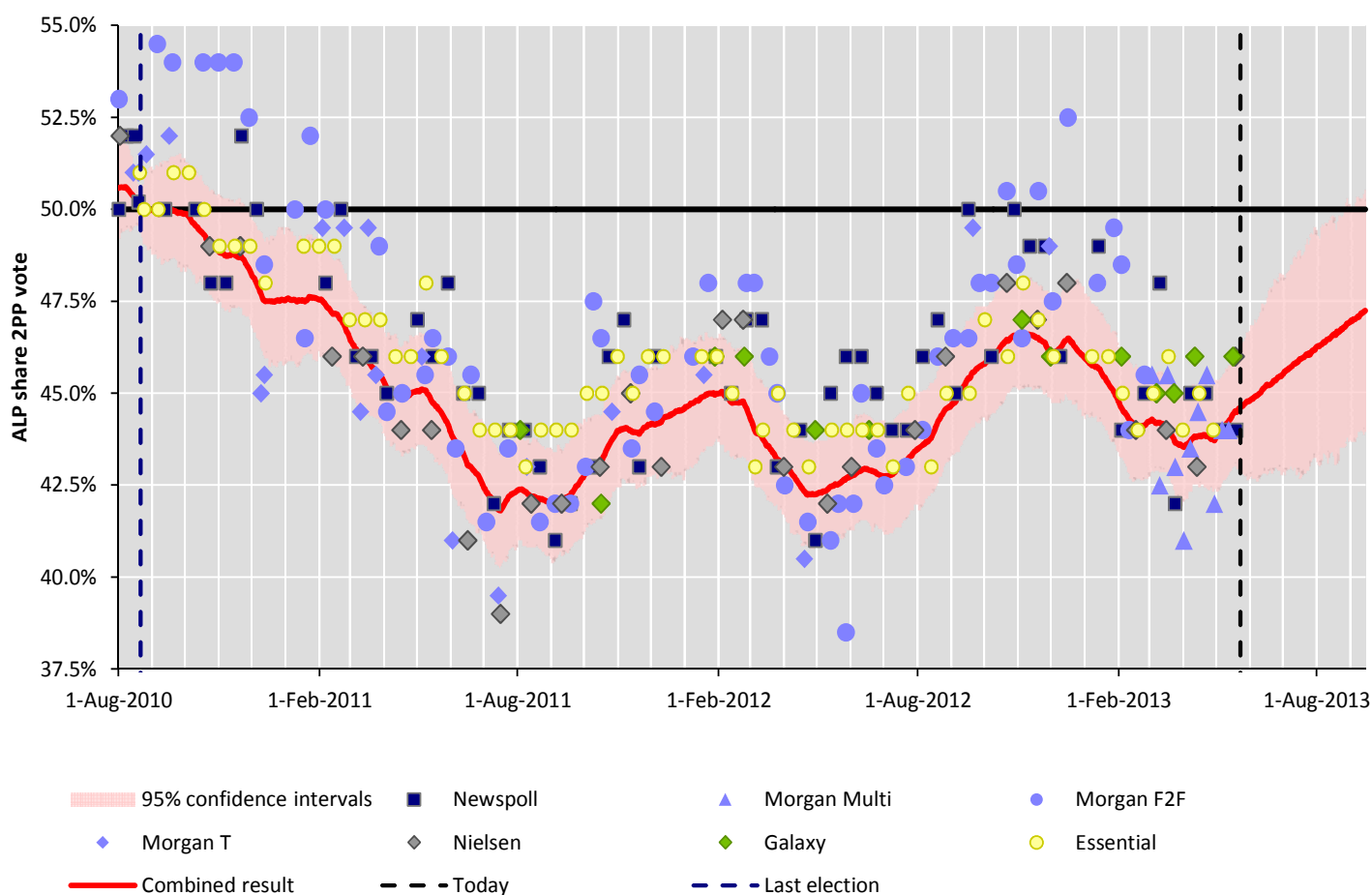
Figure 1: Historic polling data and estimated ALP share of 2PP vote between 1 August 2010 and 14 September 2013

split typically in the range 53/47. For the ALP, a 2PP result outside the range 44% to 55% is exceptionally unlikely based on elections since 1970. Using the techniques of Bayesian analysis, we incorporate this knowledge into our model.

Doing this makes our predictions more precise. That is, by using an additional independent piece of information, the uncertainty in our predictions will be lower than if we just used polling information. But, more importantly, it also makes our predictions more accurate. The ALP 2PP vote share at the moment is very low, and hovering near the bottom of the historical range (about 44.6%). We know from decades of elections that a 2PP result below 44% is very unlikely, and so our model (if it is a good one) should incorporate this fact. By incorporating this fact into our model, our model naturally predicts that the ALP's 2PP vote share will likely rise somewhat between now and election day.

Given the speed at which sentiment typically changes, it is possible that the ALP could even stage a recovery to win from this point. Given the current polling most people

would consider this unlikely, but the Coalition achieved a similar feat at the 2001 election. The ALP polled extremely well throughout the whole of the 1998-2001 electoral period and up until several months before the election. In the space of two months, the Coalition managed to achieve a reversal of fortunes to win the election with a 2PP vote share of 50.95% after the combination of the Tampa incident and the September 11 terrorist attacks.

Betting markets

Besides polls, the other real-time and observable measure of electoral sentiment relates to the betting markets. It is believed that betting markets may be better predictors of election outcomes than polls. There are a number of reasons for this, but the two strongest reasons are that people who bet on the outcome of an election are financially invested in their decision, and therefore will work hard to predict the correct outcome, and that the betting markets should make allowance for all available information (more on this in a future issue). Observationally, betting markets in the US appear to be

somewhat better at predicting electoral outcomes than polls (and in any event are no worse).

Centrebet has operated a betting market on the outcome of the four Australian federal elections since 2001. The betting odds give you an implied probability of each party winning. The implied probability of a win can be compared to the actual 2PP vote share, and we can then model the statistical relationship between these two factors. Although there are only a small number of data points for the betting market and Australian federal elections (one for each of the four elections since 2001), the predictive power of election-eve odds is excellent, with a typical 2PP prediction error of only around 0.6%.

Some betting market participants may have access to "inside" information, such as internal polling conducted by political parties. Given that the betting markets in Australia are not very deep, a few insiders might have a very significant impact on the odds.

We have modelled the relationship between the betting market odds and the actual 2PP electoral outcome, taking into
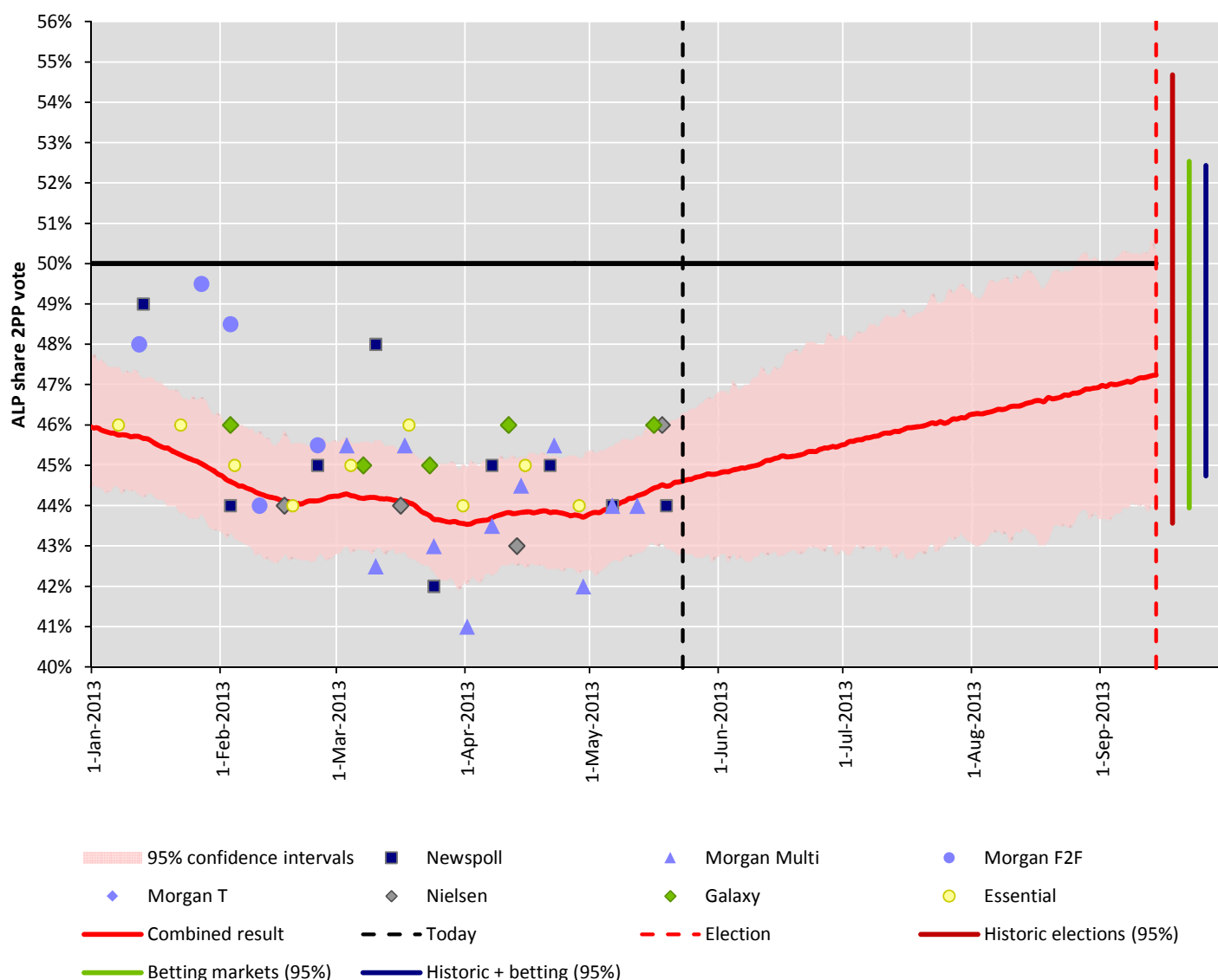
Figure 2: Historic polling data and estimated ALP share of 2PP vote between 1 January 2013 and 14 September 2013, and constraints on election day voting intentions from historic elections and Centrebet

account how much time is left to the election. Using the current betting market odds, our model predicts the likely 2PP vote share on election day and then combines this prediction with the prediction from the polling data as well as constraint from previous elections.

Our model links each adjacent day through the model for how fast voting preferences change through time. As a result of this, the betting market prediction for election day affects the estimated 2PP vote share on days before election day.

The outcome

Figure 1 shows our model for the ALP 2PP vote share from the time of the previous election until the date of the next election (14 September 2013). The different points represent different polling results from

different organisations. The red line shows our best estimate of the ALP 2PP vote share at different points in time.

Our model produces a prediction of what the 2PP vote share is likely to be in the future, on each day through to the election. The value of the model at each day in the future is the median value from our simulation-based model given our current understanding of the 2PP vote share (based on polls) and the assumptions we have made about the outcome on the election day (based on the betting markets and our prior information about what election outcomes are typical).

Figure 2 shows a zoomed in version of this graph, from 1 January 2013 until the election day. To show the impact of the information from previous elections and the betting markets, shown to the right of the election day are lines which represent the 95% range

on the prediction for the result from: historic elections (dark red bar), the betting markets (green bar) and the combined result from historic elections and the betting markets (blue bar).

Our model currently predicts a central 2PP outcome for the ALP of 47.2%, with a 95% confidence interval of about 43.8% to 50.2%

What about the number of seats?

2PP vote share is a very good determinant of the electoral outcome, although it is possible to still form government with a 2PP vote share of slightly less than 50%.

It is possible to turn predicted 2PP vote share into predictions about the number of seats won, but this requires assumptions about how a national 2PP share translates into the results in individual seats. The simplest assumption used is one of a uniform
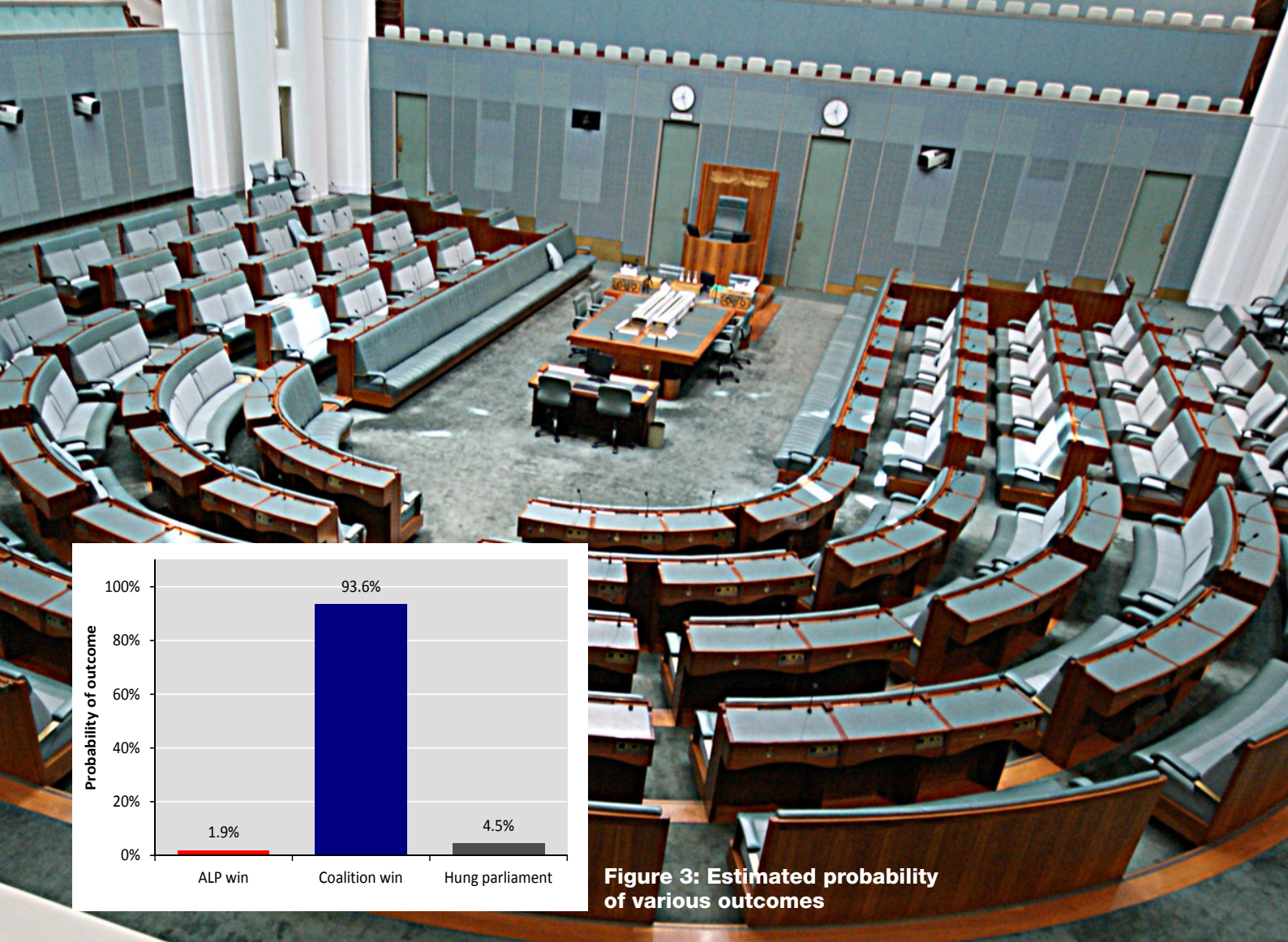
Figure 3: Estimated probability of various outcomes

swing. We have applied the predicted 2PP outcomes to the electoral results from the last election, with the electoral information taken from Antony Green's excellent election calculator. Additional assumptions are required around what happens to those seats currently held by independents. Of these five seats, we assume two seats go to the ALP, two go to the Coalition and one remains independent.

Figure 3 shows our predicted outcome for the number of seats. Based on this distribution, we conclude that the Coalition has a 93.6% chance of winning the election, the ALP has a 1.9% chance of winning the election, and there is a 4.5% chance of a hung parliament. This prediction for a Coalition win compares with the prediction for a Coalition win from the betting markets of 86.5% (Wed 22 May). Some discrepancy arises between these because the betting markets are looking at the ultimate outcome for government, whereas we have looked at the distribution of seats and made no judgement as to who the remaining independent will side with. That our numbers align well with the results from the betting markets is no surprise – at present, most of our information about the election day outcome is being driven by our model for the betting markets.

Where to now?

There is still a significant amount of time to go until the election. Although our model suggests that there is a high probability that the Coalition will win on election day, this is based on information currently to hand and is based only on national data (we'd be delighted to have access to polling data in key marginal constituencies!). The result is certainly not a foregone conclusion – the uncertainty bands on our election day result include the possibility of the ALP winning. History shows that a recovery of this magnitude is certainly possible.

We will continue to update and refine our model over the coming months, incorporating newer polling data and outcomes from the betting markets. As we get closer to the election day, the relative impact of the polling data and the betting market data will grow, with the predicted distribution of voting outcomes narrowing considerably.

In future updates, we will address a number of questions about how our model works. These include:

- The effect of polling bias – so-called "house effects"
- The issue of how good the polls and betting markets are
- Whether the betting market data and

polling data are truly independent, and the impact of this on the model (i.e. are we "double-counting" by including the polling data)

Notes

Our poll aggregation is based on the random walk with house effects model described by Simon Jackman in "Pooling the Polls over an Election Campaign" (Australian Journal of Political Science, 2005). We are grateful to Simon for data about previous elections as well as various discussions.

The temporal component of the model is implemented within VBA using standard Gibbs sampling Markov Chain Monte Carlo (MCMC) techniques. Our incorporation of betting market data and the historical data around election outcomes is carried out using standard Bayesian statistical techniques. In particular, the election day prior is created by combining the constraint from the prior elections and the betting model prediction. The constraint from previous elections is derived from a kernel density estimate of the ALP 2PP vote share from all elections since 1970. The betting prediction includes full allowance for the uncertainties in the parameters estimated. Election day samples are drawn using the Independent Metropolis-Hastings sampler with a reference function
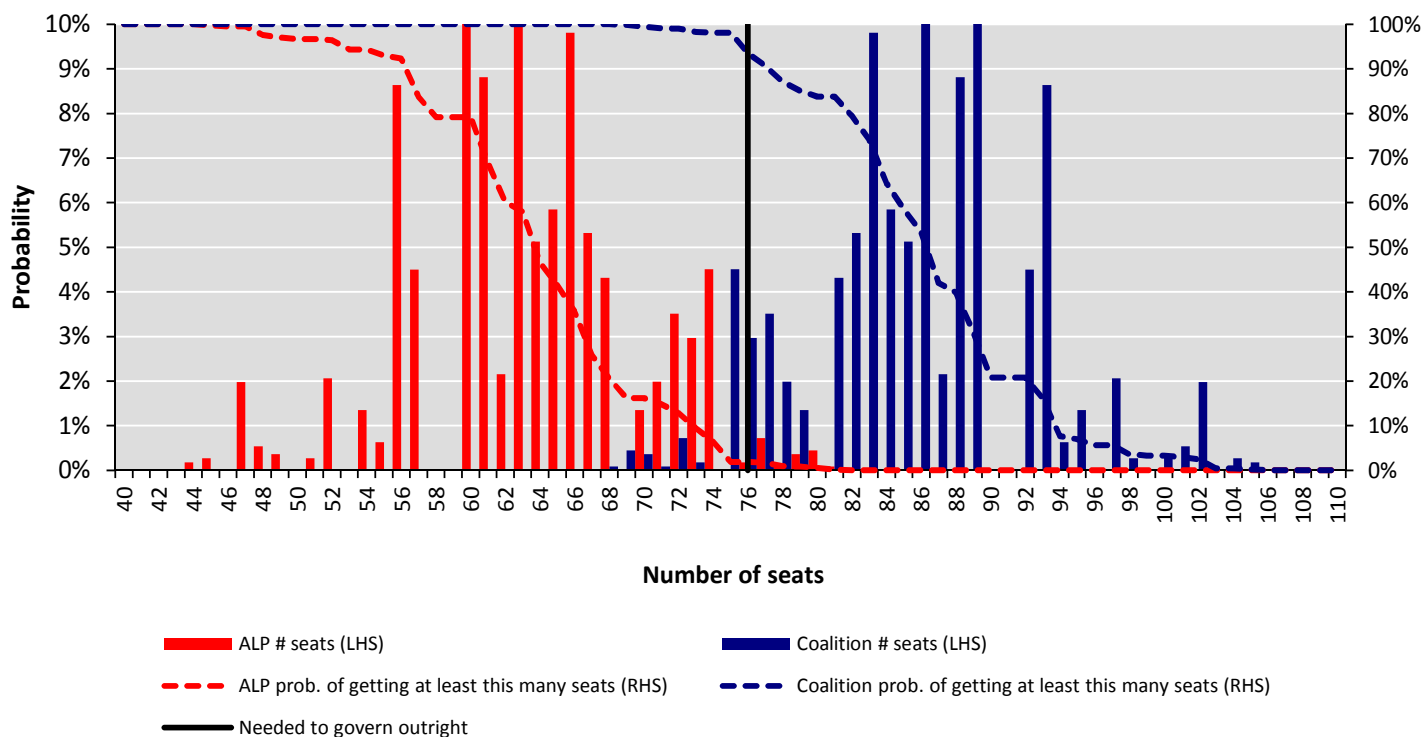
Figure 4: Implied distribution of outcomes for the number of seats for the ALP and the Coalition

that is calculated to give acceptable results in terms of mixing of the local Markov chain.

Our model also extends Jackman's model by making allowance for overdispersion of the results from various polling agencies (and their individual polling methods) about the combined 2PP estimate. Where the results from one particular class of polls are overdispersed, their standard errors are increased to compensate. In fact, this is done in a fully Bayesian way, with the overdispersion modifiers included as part of the Gibbs sampling. **P**

By Julian King

Julian is a Vice President at Pottinger. He has a PhD in astrophysics and was a winner of the 2012 Eureka Prize for Scientific Research. He has a keen interest in understanding the true drivers of risk and value in businesses, in part by trying to apply Bayesian statistics to everything.

About Pottinger

Our clients say that we offer a completely different proposition to traditional consulting and investment banking advisors, seamlessly integrating true strategic thinking, commercial insight, financial expertise and execution excellence. Our assignments typically relate to one or more of:

- Strategy and public policy
- Mergers and acquisitions
- Partnerships and joint ventures
- Restructuring and capital advice
- Risk, sustainability and related decision-making

Our approach to every assignment reflects a fundamental belief that strategy, business and execution perspectives must underpin any business initiative if it is to be commercially successful and stand the test of time.

Together our team has advised on over 200 M&A and financing transactions, as well as many significant strategic advisory assignments. Our first hand experience covers most of the world's larger economies, and we are accustomed to working on complex assignments across borders and cultures.

We are highly regarded for our investment in people, most recently being profiled by the Australian Workforce and Productivity Agency as a role model for effective skills development in financial services. In addition, Pottinger is the only organisation ever to have won the ABA's "Recommended Employer" award for six years in a row.

Cassandra Kelly
Joint CEO

Nigel Lake
Joint CEO

For further information, please contact either of our joint CEOs.

Level 35, AMP Centre
50 Bridge Street
Sydney NSW 2000
Australia

p +61 2 9225 8000
w pottinger.com

Past issues from Pottinger Perspectives:

Pottinger
PERSPECTIVES

Courage - Integrity - Excellence - Respect - Commitment - Passion • www.pottinger.com • April 2013

THE DRAGON'S BEST FRIEND
Unleashing The Potential For Growth

Fifteen years on from the Asian crisis, the contrast between the fortunes of East and West is stark. Europe's economies continue to be plagued by high unemployment, with youth unemployment in some regions now exceeding 50%. Meanwhile, previously stable countries have been forced to face the possibility of economic collapse, as the effects of the global financial crisis continue to be felt five years after its beginnings. The result: the lowest growth experienced in many years and continued uncertainty.

In contrast, the Asian region continues to grow powerfully. China's economy has expanded by more than 300% over the last decade. Even Australia's economy has grown by some 30% over that time, reflecting the benefits of exposure to both China and India, and both economies have moved up the world rankings. Looking forward, Australia has the potential for sustained growth if it can continue to harness the opportunities that China offers. A key to unlocking the potential will be for both countries to understand clearly each other's cultures and each other's needs to figure out where the most attractive areas of mutual opportunity lie.
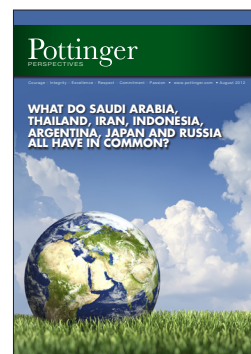
Read More...

Australia: The Asian Home Of Innovation
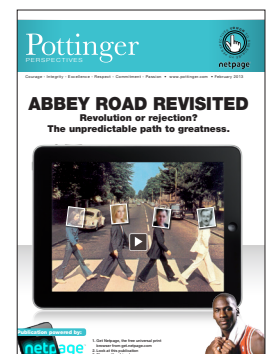*Securing Australia's future as an Asia-Pacific finance hub.*
Read more...

Hopenomics or Leadership?
Securing Australia's prosperity in an uncertain world.
Read more...

Embrace madness
Read more...

What Do Saudi Arabia, Thailand, Iran, Indonesia, Argentina, Japan And Russia All Have In Common?
Read more...

Abbey Road
Revolution or rejection? The unpredictable path to greatness.
Read more...

Please visit www.pottinger.com/think to see all our latest news and articles from the team!