

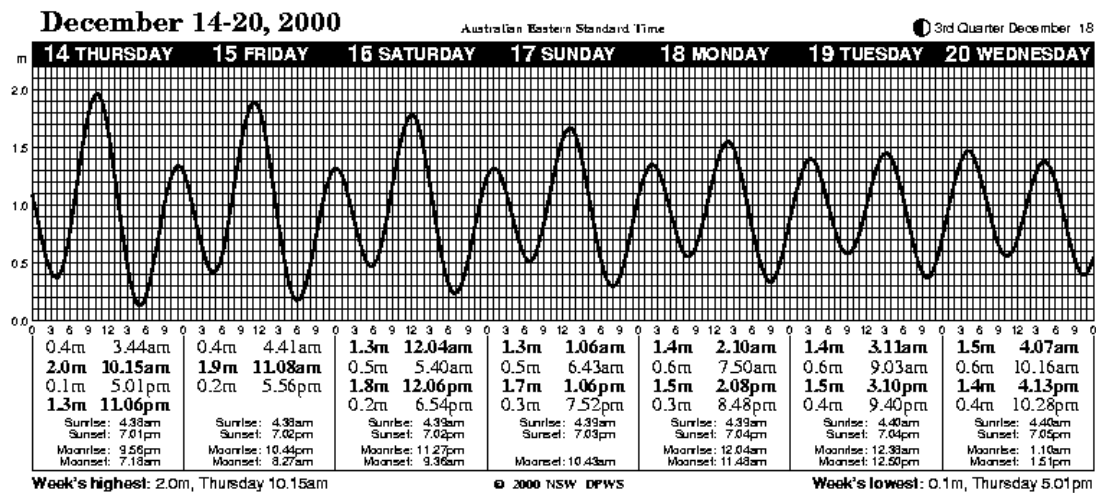
Regime Switching and Cycles

By Colin Priest BEc FIAA

Introduction

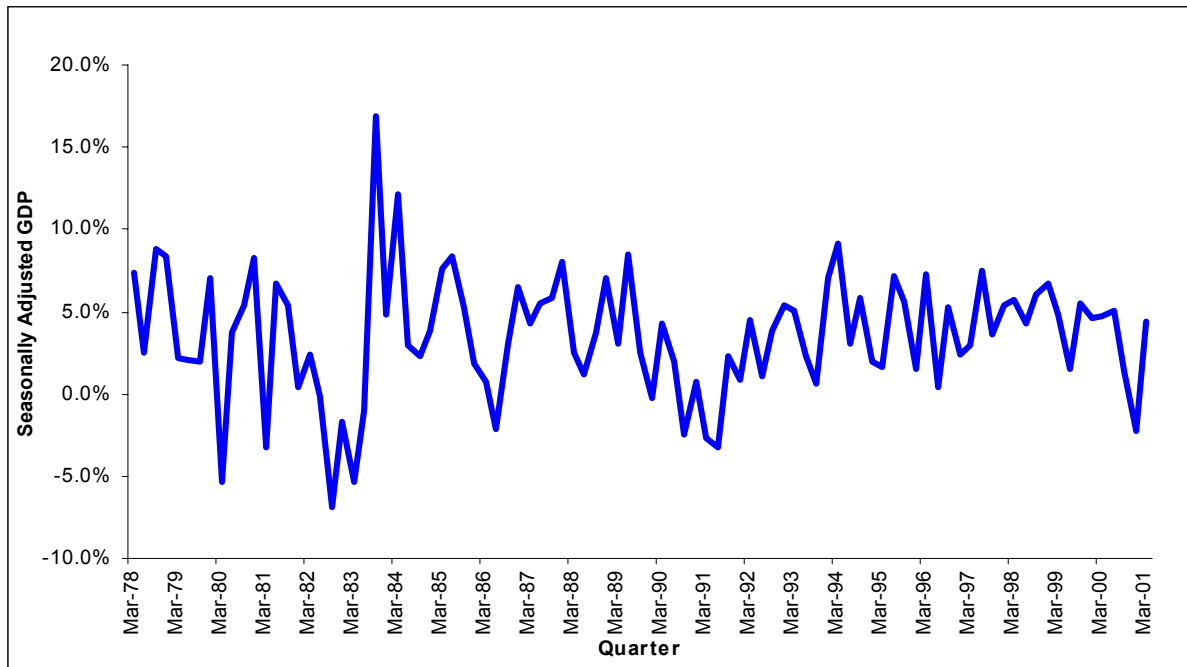
Many things in life move in cycles. Some, like the seasons and the tides, move in cycles that can be predicted perfectly using well understood laws of nature.

Figure 1 : A Tide Chart



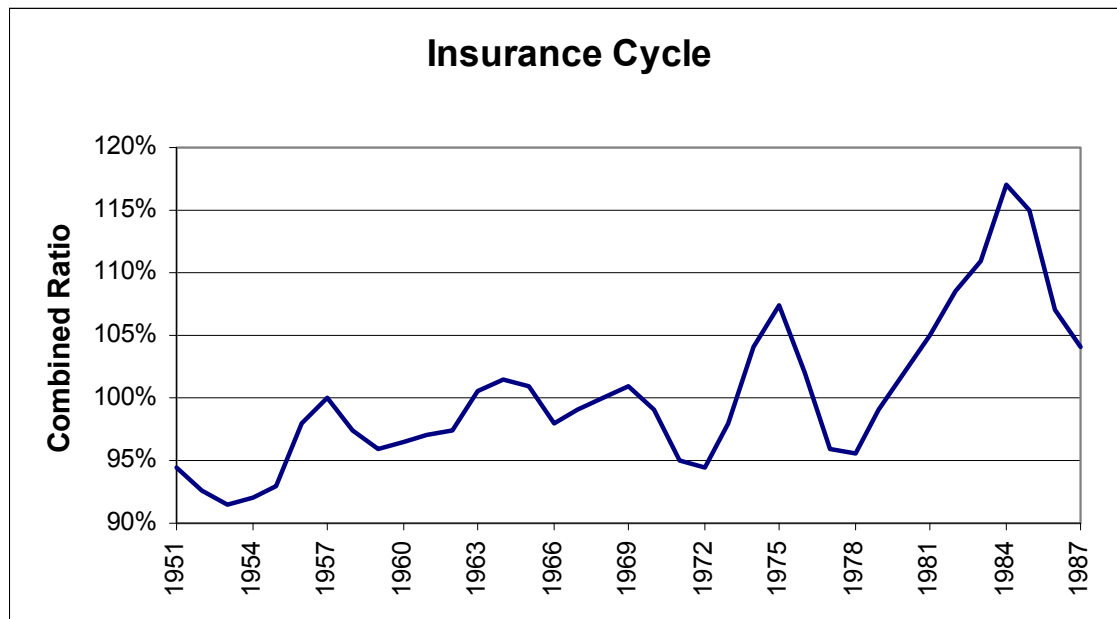
Others, like economic growth, are harder to predict because they don't follow any known rules and because each cycle is a different length and/or height to the previous.

Graph 1 : Economic Growth



Traditional statistical approaches cannot cope with these more chaotic cycles. But some of these cycles are quite important to insurance – interest rates, unemployment rates, inflation, economic growth and building construction activity are key drivers of many classes of business. The insurance cycle is itself an example of a chaotic cycle.

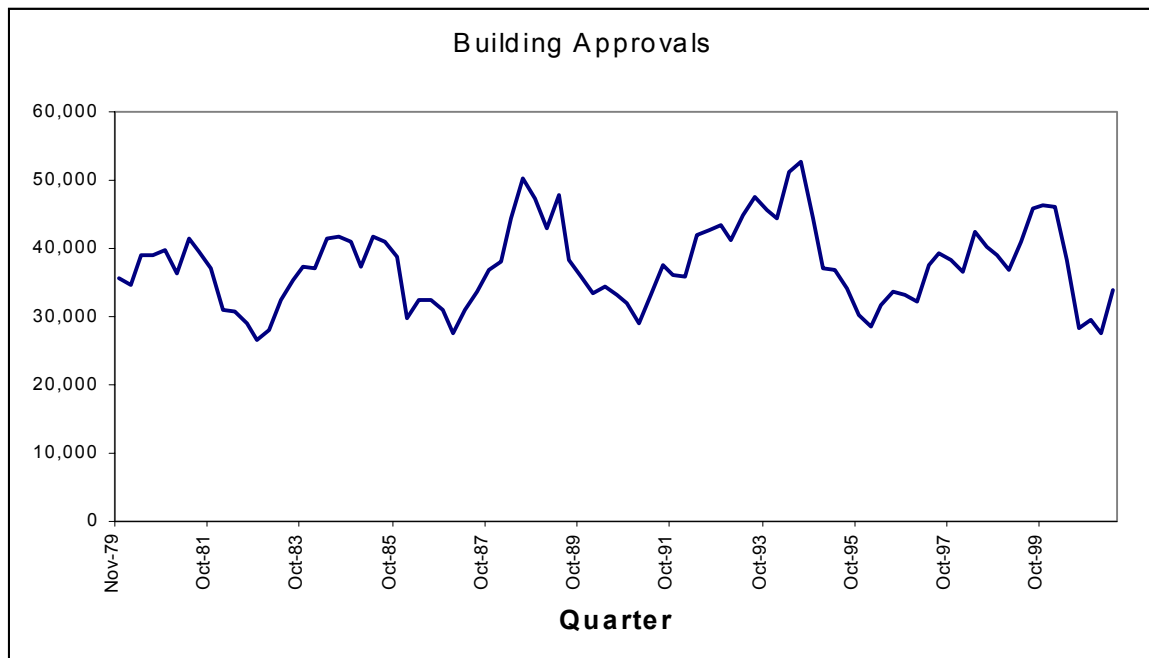
Graph 2 : An Insurance Cycle



This paper develops a new approach to chaotic cycles that allows us to forecast future values from chaotic data. It takes the reader through a case study of designing and using a model for building approval levels in Australia, and then briefly looks at how the same technique could be used to model the insurance cycle and to model changes in unemployment rates.

The Building Cycle

Graph 3 : Quarterly Building Approvals in Australia



Building approval activity in Australia has historically moved in cycles. These cycles are of semi-regular length. There is a seasonal effect, with the March quarter typically being lower than the other three quarters due to summer holiday breaks.

Due to the introduction of GST the last cycle ended remarkably quickly, helping to push Australian GDP growth into negative values.

The building cycle has three characteristics that define the sort of cycles that this paper addresses:

1. A floor and a ceiling for values: the turn-around points for the peaks and troughs occur around similar heights
2. A fairly constant slope for each movement between a particular peak and trough: the lines joining each peak and trough are fairly straight

3. A semi-regular cycle length: cycle lengths do vary, but they are similar to another

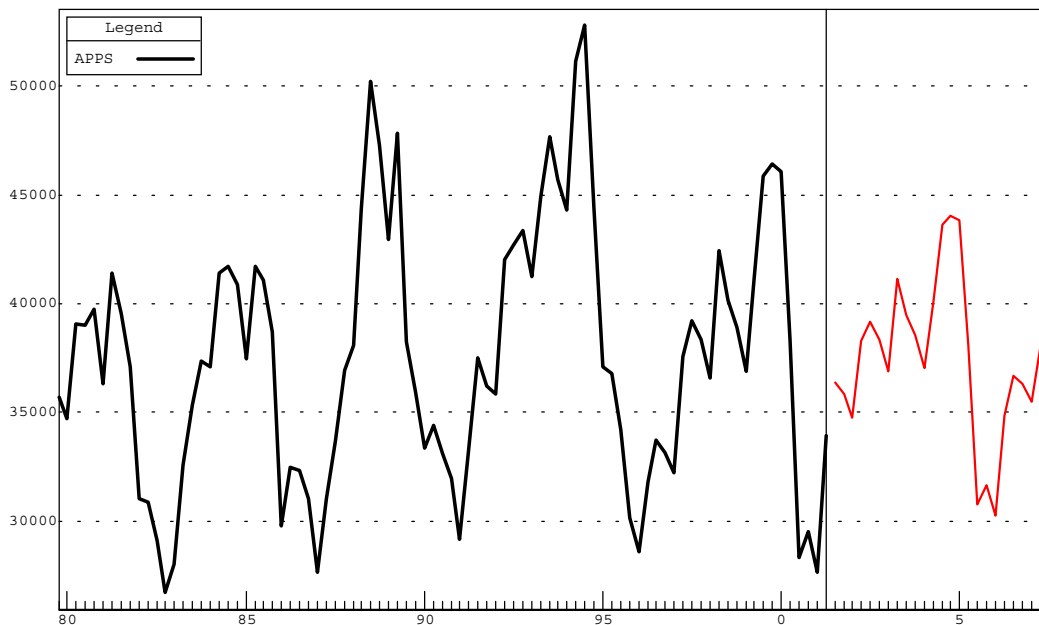
The Failure of Traditional Forecasting

Some time ago I purchased some specialist statistical forecasting software. The software has an "expert selection" option that attempts to fit a range of models including exponential smoothing and a Box Jenkins approach to traditional time series.

Unfortunately the software was not enough of an expert to determine the length of a cycle and relied upon the user to tell it the length of a cycle period before it would fit a model. Having to tell the software the length of a cycle was the opposite of what would normally be wanted – we usually want to know what the data infers about the length of cycles.

This approach also has the drawback that each cycle is fixed to be exactly the same length. The historical data is not consistent with this assumption. The last cycle dropped far more quickly than in any previous cycle, and the upward section of the last two cycles lasted longer than those of the previous cycles.

Graph 4 : Forecast Using Traditional Methods

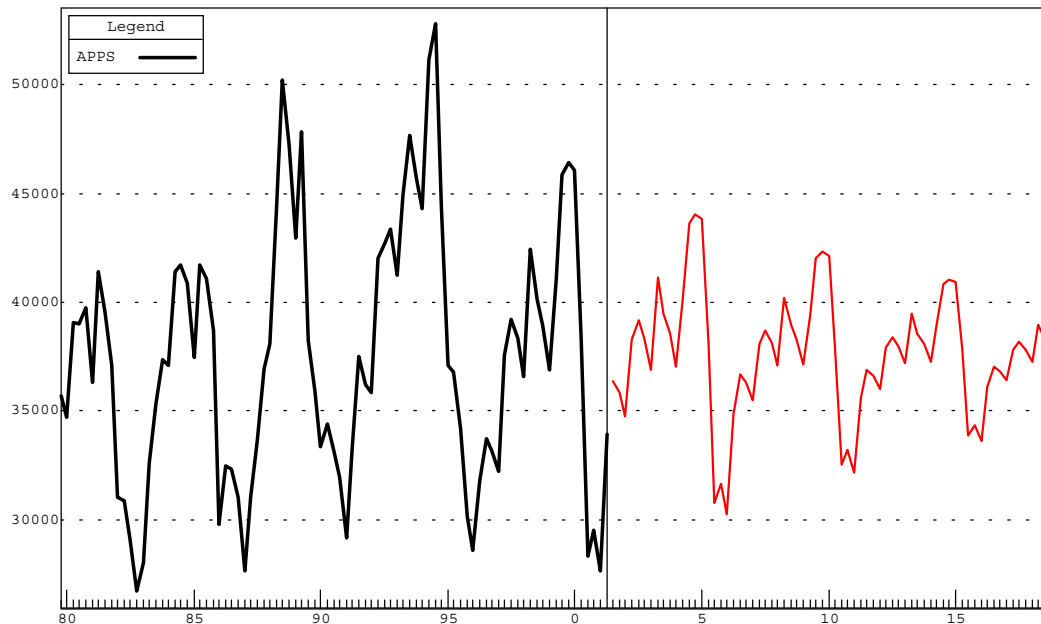


The best forecast that came out of the system looked reasonable at first. However it does suffer from the problem that it is too much like the last cycle in

shape, and we know that the last drop in building approvals was unusual and unlikely to be repeated because it was due to one-off changes to tax.

The forecast performed even worse once it was extended beyond one cycle.

Graph 5 : Longer Forecast using Traditional Methods



The forecast cycles have decreasing amplitudes! With the benefit of the extra forecast periods, we can see more clearly that even the first forecast cycle suffers from decreased amplitude (the distance in height between the top of the cycle and the bottom of the cycle). Dampening of cycle amplitudes is a feature of mean reversionary models, and is caused by autocorrelation factors less than 100%.

If the purpose of the forecast were to price some form of stop loss reinsurance, then the forecast is totally inappropriate. It underestimates the height of the peaks and troughs of the cycle, and it does not allow for any variability in the length of cycles.

What is a Regime?

For the purposes of this paper, a regime is a time period in which a particular set of rules apply. Once the regime is over those rules no longer apply, and another set of rules, belonging to another regime, come into force.

A real life example of regimes is a tennis match. There are four distinct regimes:

1. Player A is at the south end of the court and is serving
2. Player A is at the south end of the court and is receiving
3. Player A is at the north end of the court and is serving
4. Player A is at the north end of the court and is receiving

The rules for each regime vary, changing which direction in which Player A must hit the ball (either north or south) and whether Player A begins the play (whether Player A serves).

Player D, had a regime switch in tennis when she changed nationalities.

Figure 2 : Player D (or is her father the real Player D?)



Regime Switching Models

Regime switching models are characterised by a number of distinct and discrete regimes within which different model parameters apply. From time to time the model switches from one regime to another and the characteristics of the observations change to match the underlying characteristics of that particular regime. These regime switches represent structural changes occurring in the process being modelled.

The use of regime switching models began in the field of economics. Kim and Nelson (1996) describe the development of regime switching models within econometrics. Of particular interest is the use of regime switching models to describe economic cycles in which economic growth and recession are the two possible regimes. An economy can be seen as cycling between boom and bust.

A regime switching model allows the econometrician to measure the different characteristics of growth and recession.

Regime switching models are not unknown to actuaries, especially here in Australia. Harris (1999) describes an actuarial application of regime switching models in investment modelling. He derives a regime switching vector autoregression model that describes investment returns for different classes of assets. His model has different regimes for stability and uncertainty.

The reason that regime switching models have only come into use in relatively recent times is that it is difficult to fit parameters for regime switching models using traditional least squares and maximum likelihood methods. At no point of time can one directly observe which regime a process lies in. Regime choices can only be inferred by observed data. So the regime values must be treated as unobserved data.

Bayesian Statistics

There is a branch of statistical theory called Bayesian statistics, named after the Reverend Thomas Bayes. Bayes lived from 1702 to 1761 in England and was an amateur mathematician who first proposed a mathematical basis for probability inference (the way that estimates of probabilities are affected by observed data and by certain events occurring).

Figure 3 : Rev Thomas Bayes



Bayes is most famous for the theory named after him. Bayes Theory is:

$$P(AB) = P(A | B) * P(B)$$

Or one of wording it in English is that the probability that events A and B will both occur equals the probability that A occurs conditional upon B having definitely occurred, times the probability that B occurs.

This theory is important because it forms the basis of some of the mathematics used to fit parameters for regime switching models. It introduces the concept of a conditional probability.

To understand a conditional probability let's consider a hypothetical general insurance seminar attendee called John. John has a drink in the evening with other people from the general insurance seminar in two of the three evenings, but he does not drink during the daytime hours. If we let a computer select a random point of time in the seminar, what is the probability that John is drinking at that chosen time?

We know the *conditional* probabilities of John drinking. If it is during the day, then the probability is zero. If it is the evening, then the probability is two thirds. These are conditional probabilities, because they depend upon the status of another variable, in this case the time of day.

Bayes theory shows us how to combine these probabilities. There is a 50% chance that it is during the day and a 50% chance that it is during the evening. The probability that John is drinking and it is the evening is the probability that John is drinking conditional upon it being the evening times the probability that it is the evening. The answer is a probability of one third.

Judging by the stern look on the portrait of Rev Bayes, I suspect that the probability of finding him drinking was zero!

The technique that is used to fit parameters to cycle models in this paper is just a fancy version of the example just presented.

Gibbs Sampling

There's a little bit of algebra here. But don't worry – you don't need to know the algebra to follow the remainder of the paper. You can skip the next four paragraphs if you're not a technophile, and go straight to the crossword analogy.

Gibbs sampling is a Markov Chain Monte Carlo algorithm for approximating joint statistical distributions by sampling from conditional statistical distributions. It uses a Bayesian approach – so it uses probabilities that are conditional upon some values being certain or fixed, and combines these probabilities so that all of the values are variable.

Suppose you want to estimate a multivariate probability function $f(x_1, x_2, \dots, x_N)$ but that doing so directly is computationally unfeasible. The Gibbs sampler algorithm can produce samples from this multivariate distribution using the following algorithm:

1. Start with an arbitrary set of starting values (z_1, z_2, \dots, z_N)
2. Draw a sample from $f(z_1 | z_2, \dots, z_N)$
3. Draw a sample from $f(z_2 | z_1, z_3, \dots, z_N)$
4. Iterate through j (where steps 2 & 3 are the first two iterations) to sample from $f(z_j | z_1, z_2, \dots, z_{j-1}, z_{j+1}, \dots, z_N)$
5. The last iteration of j is $f(z_N | z_1, z_2, \dots, z_{N-1})$

Steps 2 to 5 can be repeated a number of times. The distribution of the set of sample values stored at step 5 will converge to $f(x_1, x_2, \dots, x_N)$.

You can store the sample values and use them to approximate $f(x_1, x_2, \dots, x_N)$. But you wouldn't store the early sample values because it takes a while for the sampling series to converge to the target distribution when the starting values are arbitrary. In my experience, the starting values cannot be totally arbitrary because it can prevent the series from converging at all.

Crossword Puzzles

Gibbs sampling is a little bit like solving crossword puzzles. When I do a crossword puzzle, I usually do the easy clues first because they can provide hints about how to do the difficult clues.

How difficult would it be to solve each clue of the crossword puzzle without knowing how the answer fit into the puzzle?

Some of the questions would be easier than others:

12 Across *Oddly buoyant craft* (4)

In this case the word "oddly" means to take the odd numbered letters in the next word, spelling out "boat". The definition is "craft". It's pretty clear that the answer is "boat".

Some of the questions have me stumped:

1 Across *Cornetist's panic upset other musicians* (7, 8)

Some of the questions are such that I can come up with a few different answers:

25 Across *Take charge of an old horse inside? I object!*

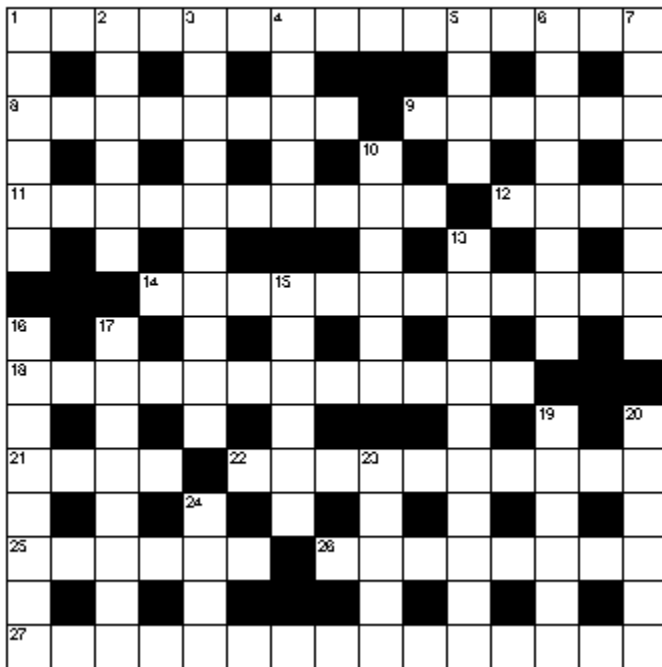
The answer could be "amount". A horse could be "a mount". "I" could be "a". A charge could also be an amount.

The answer could be "manage". "An" could be "a". An old horse could be a "nag". "I object" could be "me". "Inside" could be an instruction to place the first two answers inside the last, forming "manage".

Without knowing anything of the words that cross my answer, I cannot determine the answer. So what I need to do is look at some of the words going down.

Figure 4 : A Cryptic Crossword

Cryptic Crossword 16342 Sydney Morning Herald



From the crossword pattern I can see that I should look at 16, 17 and 24 down.

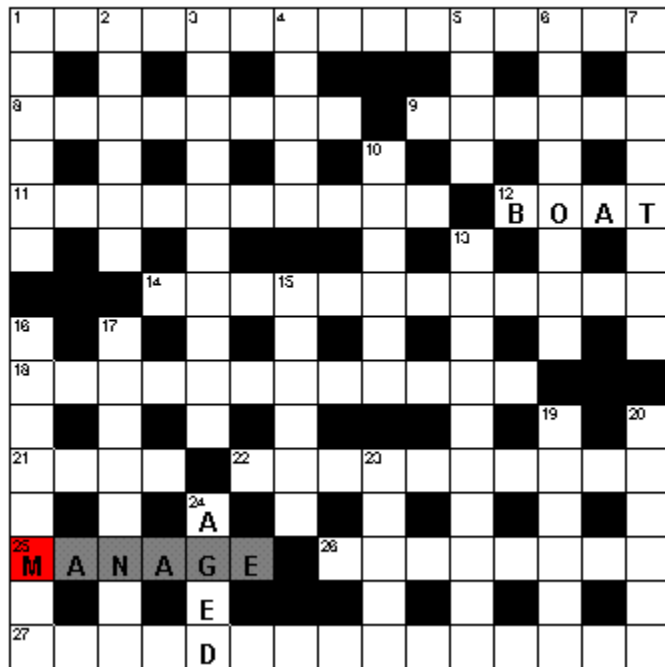
So we first look at 24 down *having lived for very many years in an average domestic situation*.

"Aged" could be an answer. The definition is "having lived for very many years". "In" means that the answer is hidden within the following phrase and the letters "aged" come from the last 3 letters of "average" and the first letter of "domestic".

This helps us with 25 across, because it means that the second last letter of 25 cross must be a "g" and only one of our possible answers fits.

Figure 5 : Partially Completed Crossword

Cryptic Crossword 16342 Sydney Morning Herald



And so we continue to work our way through the crossword. Solving one clue at a time, but using the answers from other clues to help us.

From Crosswords to Cycles

Modelling cycles using regime switching models is a bit like doing the crossword clues without the benefit of the crossword structure – you might be able to do it, but it's too difficult to be any fun. So, we use the Gibbs Sampler algorithm to split the process into easier problems that rely on the interactions between values. In our case, we separate the two key types of values:

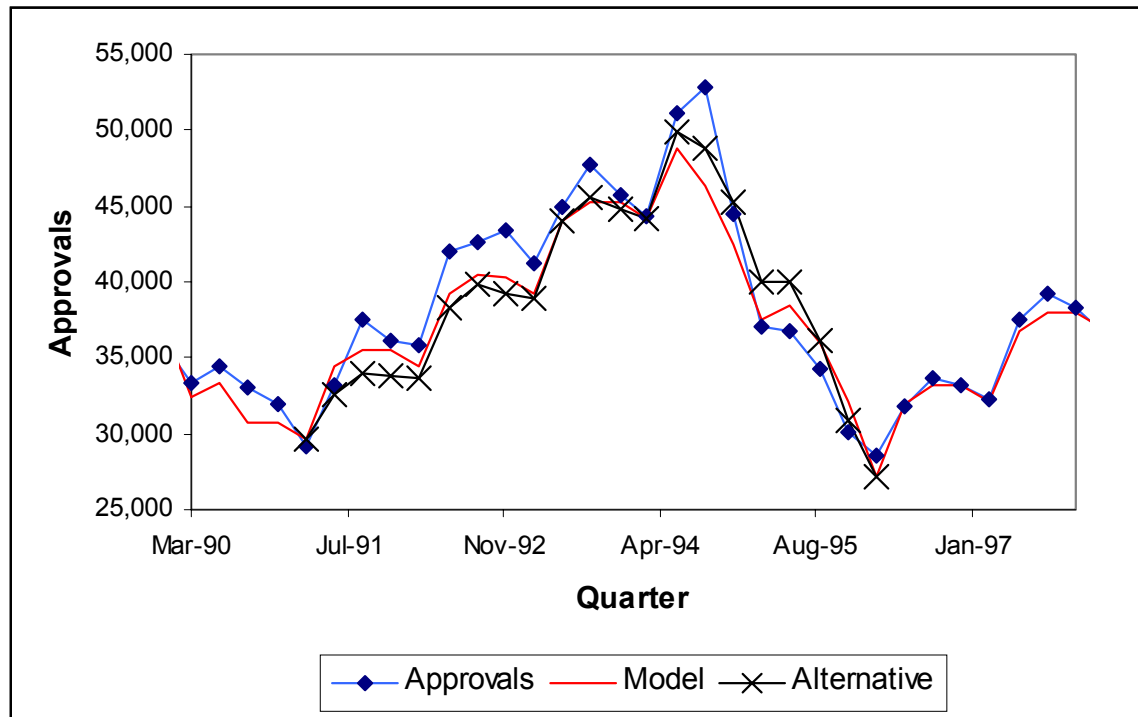
- The location in time of peaks and troughs
- The heights of peaks and troughs, and the seasonal factors within the calendar year

We deal with each turning point individually, acting as if the location of every other turning point is exactly known. Also, we assume that we know the exact height of the cycle and the exact seasonal factors. We can then look at the consequences of moving the turning point backwards or forwards in time.

In the graph below, we look at an alternative location that is one quarter later than the location that had previously been used. In practice we look at each and

every possible location (i.e. every location between the previous turning point and the following turning point) and assign a probability to each based upon the goodness of fit and upon any preconceived ideas of what the probabilities should look like.

Graph 6 : An Alternative Turning Point Location



Once we have assigned the probability for each possible turning point location, then we randomly select one of those locations according to those probabilities, and we set the location to be equal to that value while we review and set values for the locations of each of the other turning points.

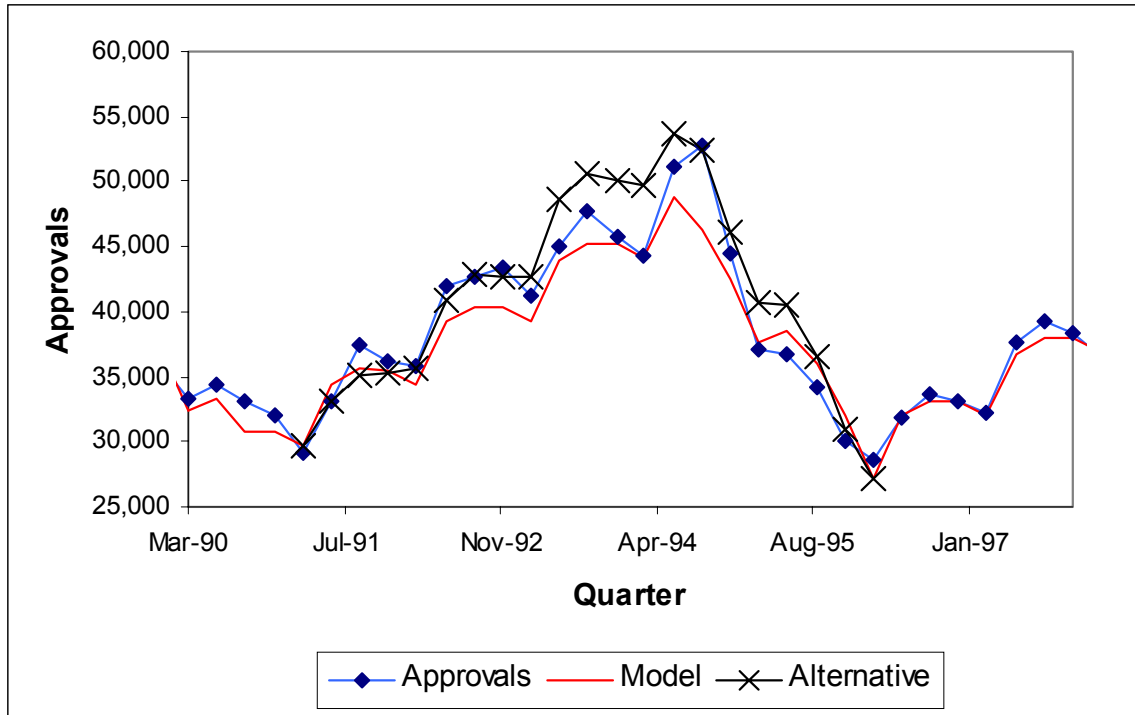
It may seem strange that random selection of values is the way to fit parameters to a model, but it works because we have constrained the sampling to be related to the goodness of fit based probability.

Once we have selected the locations of each and every turning point, then we need to set some values for the heights and the seasonal factors. We treat the problem as if the values for the turning point locations are exactly known. This simplifies the mathematics to the point that it becomes a simple multiple linear regression, although we can also allow for some preconceived ideas in the mathematics. For example, we can constrain the values so that they are never negative (how can you have a negative approval count?).

There are a number of different possible values for the heights and seasonal factors. Once again, we assign probabilities (to be pedantic, we assign

likelihoods based upon the linear regression and any prior expectations) to them, and we randomly select some of those values and assign them to be the values that we will use for the next round.

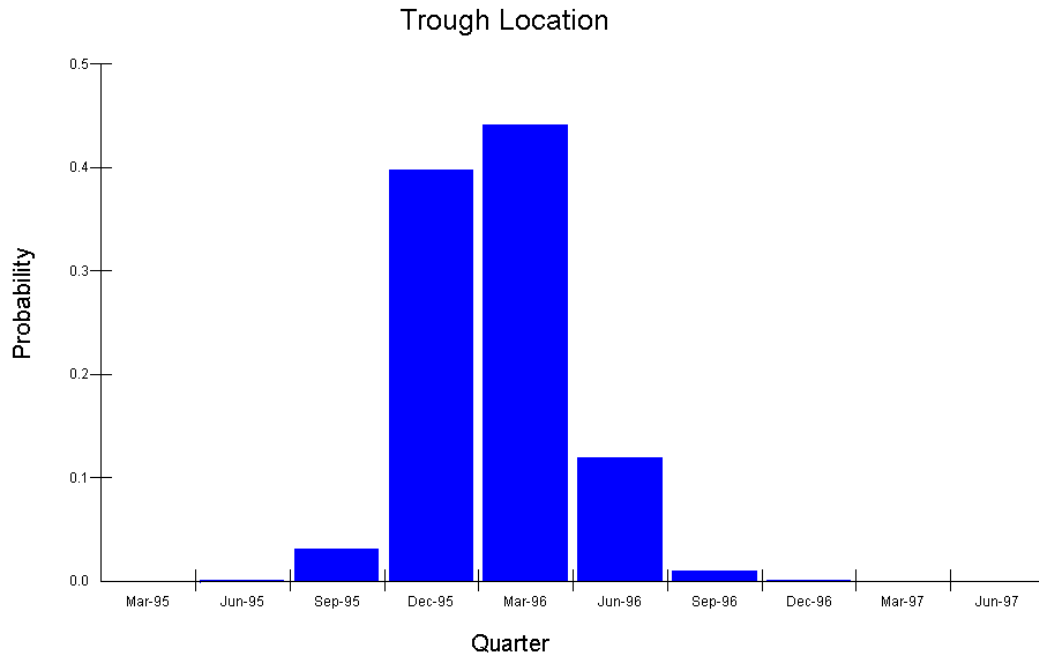
Graph 7 : An Alternative Cycle Height



Now that we have changed the height of the cycle and seasonality, the probabilities of different turning points will be different. So we repeat the turning point location steps using the new heights. This will then change the likelihoods of different cycle heights, and so on.

After each set of turning point locations, cycle heights and seasonal factors are sampled, the set of sample parameter values is stored away. Once the process has been repeated enough times, the sample values can be used to show the average parameter values and the spread of possible parameter values.

Graph 8 : Sample Histogram of A Particular Parameter Value

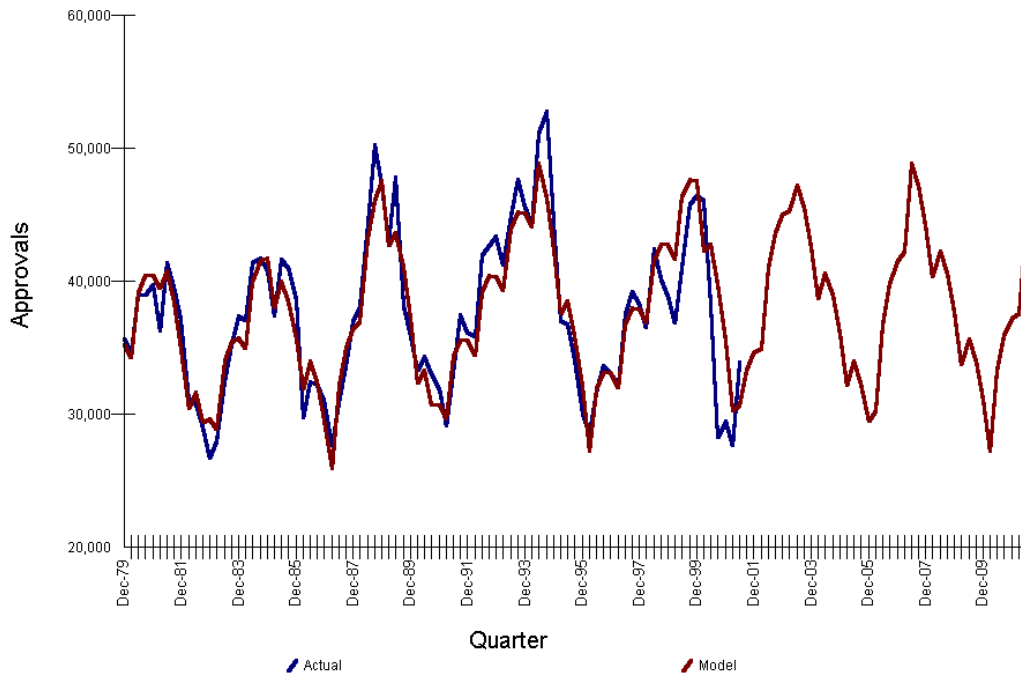


Results for Building Cycles

When this technique is applied to building approval data in Australia, we get a number of different model values, representing the range of model parameters that could be applied to the data. I will use the average values.

These values can also be extrapolated into the future.

Graph 9 : Modelled Building Approvals

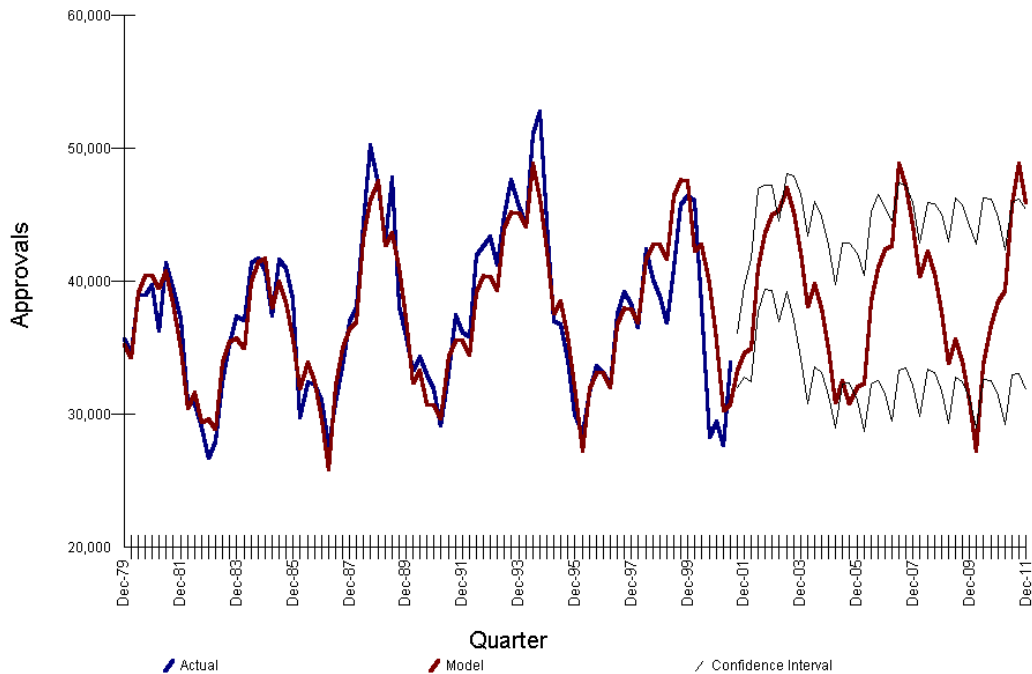


The extraordinary drop in building approvals in 2000 (due to the introduction of GST) causes the model some problems. That is because the model is an oversimplification of reality. It forms a zig zag with a seasonal factors, whereas in reality there can be periods in which the momentum of the cycle temporarily slows down or speeds up.

The forecast values are a considerable improvement over traditional time series approaches. They have a sensible amplitude, and they don't just repeat the shape of the last cycle.

One of the strengths of using the Gibbs Sampler algorithm is that we can go beyond the simple forecast and include confidence intervals.

Graph 10 : Modelled Building Approvals with Confidence Intervals



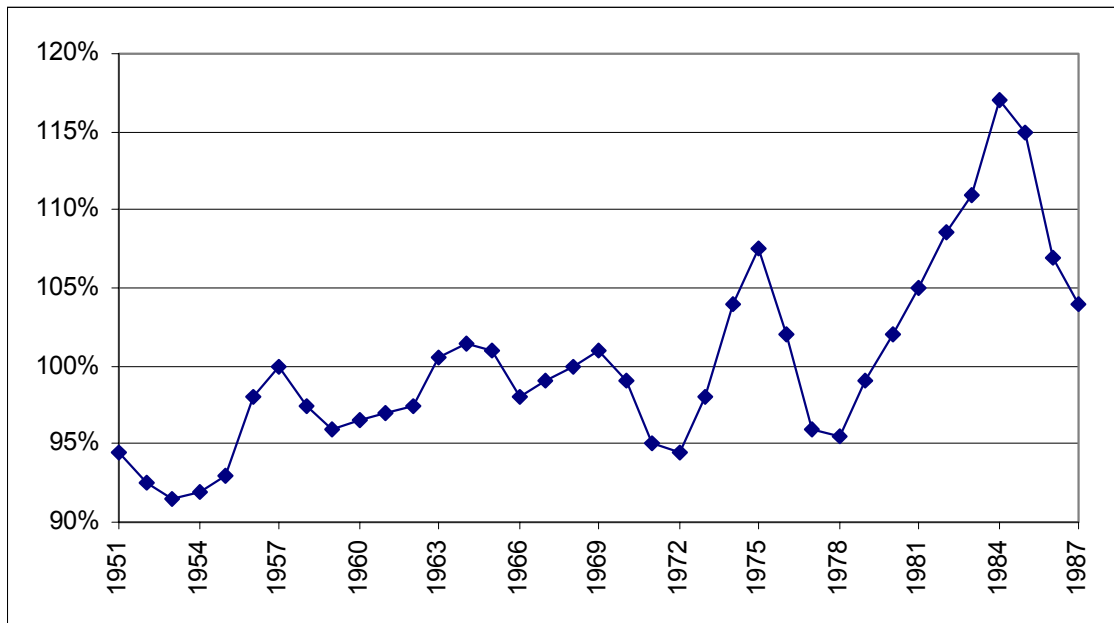
Forecasting cycles is much like forecasting the weather. We can be reasonably sure of the values for a fixed time, and then the forecasting dissolves into chaos. With weather forecasting the accuracy is about three days. For building approval cycle forecasting the accuracy seems to be about one cycle ahead.

The greatest forecasting accuracy is for the next half cycle because the model uses the data since the last turning point to narrow down the possibilities.

Insurance Cycles

Cummins et al (1991) present a history of combined ratios in the US, summarised in the graph below. We can apply the same techniques described previously in this paper to obtain an estimate of the statistical distributions of the time taken from peak to trough and from trough to peak.

Graph 11 : US Combined Ratios



A feature of the US data is that there appears to be a very regular cycle. We can use the regime switching model to measure how regular the insurance cycle has been.

The starting year of the data (1951) and the last year of the data (1987) are both allocated nodes that are held fixed, much like the starting point in the building cycle example. These two nodes are neither peaks nor troughs. They just exist at the locations at which the data has been censored.

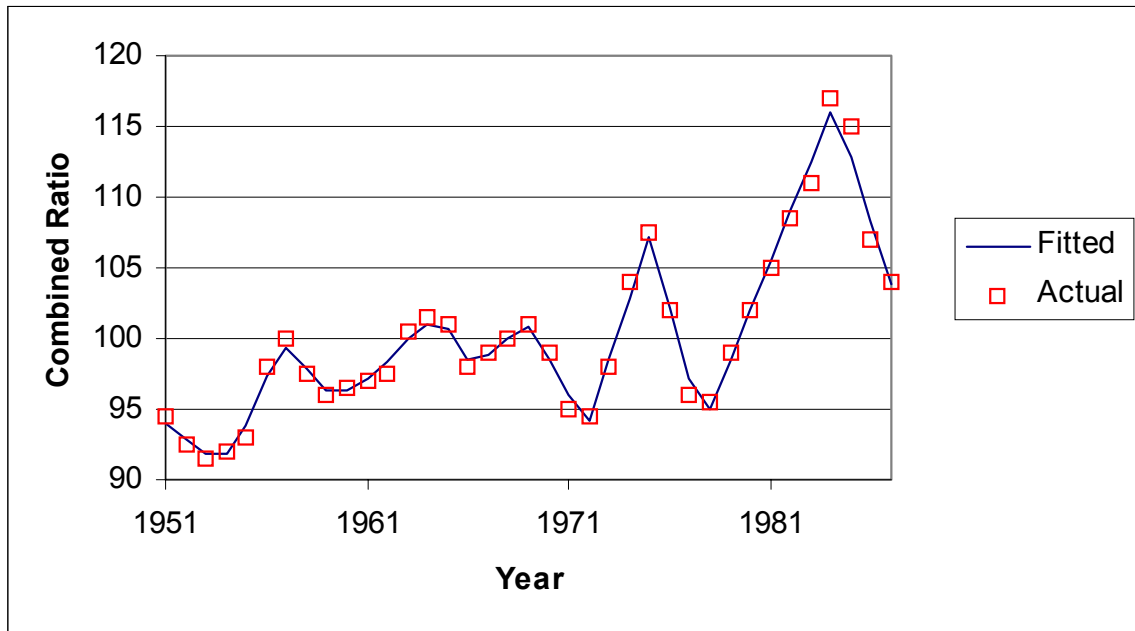
The node locations and heights are then generated in the same manner as for the building approvals data. In this example, we are interested in the distance between turning points, but the height of the nodes is required in order to generate the location of the nodes.

Table 1 : Location of Turning Points for US Combined Ratios

	Mean	Std Dev	Median	Mode
Starting Point	1951.0	0.000	0	0
Trough	1953.9	0.782	3	3
Peak	1956.8	0.482	6	6
Trough	1959.8	1.009	9	8
Peak	1964.3	0.672	13	13
Trough	1966.5	0.649	15	15
Peak	1968.8	0.482	18	18
Trough	1971.9	0.267	21	21
Peak	1975.0	0.079	24	24
Trough	1977.7	0.447	27	27
Peak	1984.1	0.352	33	33
Ending Point	1987.0	0.000	36	36

Note that the standard deviation of the estimates varies between turning points. The highest standard deviation is for the fourth node, when the combined ratio bottomed out and did not change much for a few years.

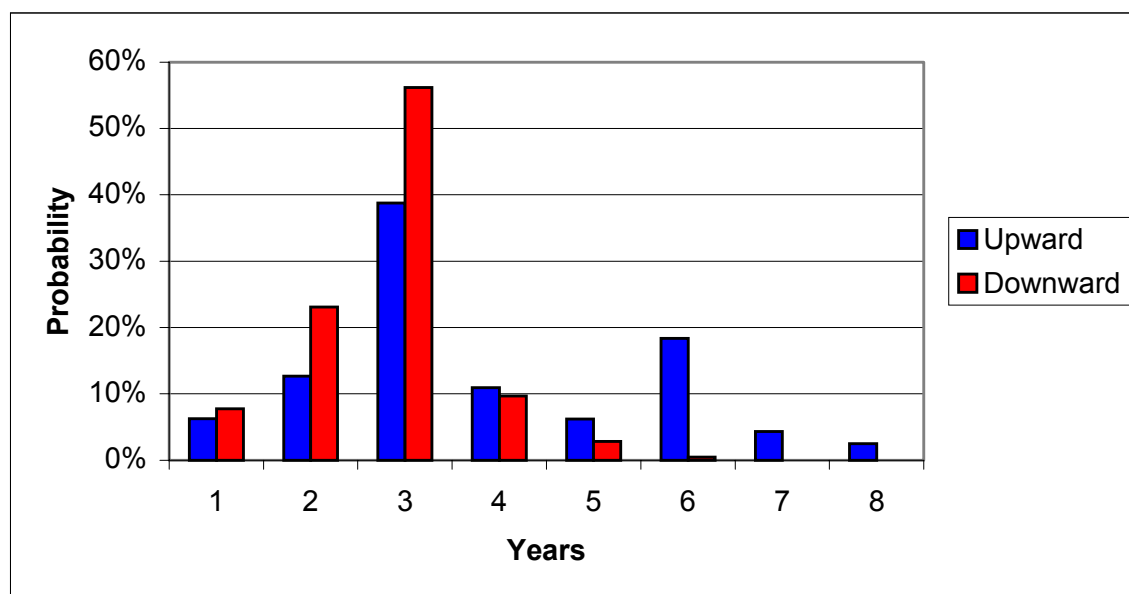
Graph 12 : Fitted Model vs US Combined Ratio History



Despite the fact that I'm only fitting a simple saw-tooth model, the model provides a good explanation of the shape of the cycle.

On to the question of how regular the insurance cycle has been, we can look at a histogram of the time gaps between turning points.

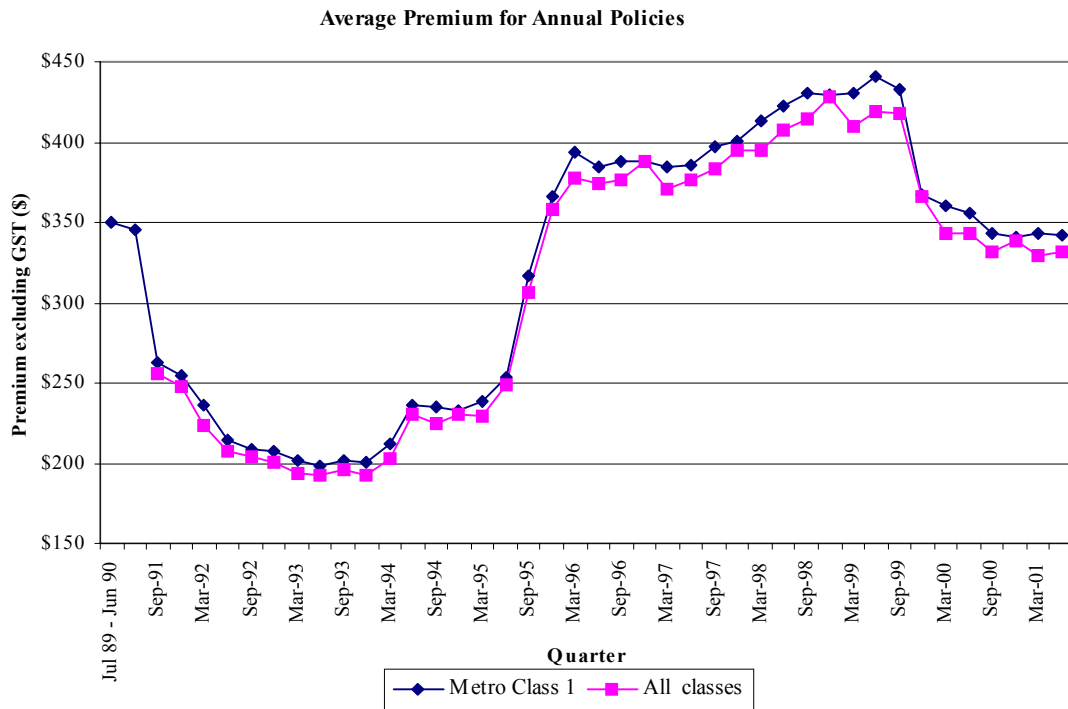
Graph 13 : Distribution of Period Lengths Between Turning Points



The most common time periods for both upward trends and downward trends is a value of three years, a value which is far more dominant than could be explained by a simple statistical distribution, such as the Poisson. This means that the timing of insurance cycles in the US has been quite regular and predictable.

Insurance cycles don't just occur in combined ratios. They can be seen in premium rates. The Motor Accidents Authority in NSW keeps statistics on the average premiums charged for NSW Compulsory Third Party policies. As a time series it shows clear signs of being cyclic. There are definite regime switches clearly visible in the graph as the legislation is changed, and as insurers follow the discovery that they are losing money and increase their rates. There is a very real upper limit to average premiums, driven by political forces. The NSW government has stated its desire to keep average premiums under \$500. This triggers a regime switch whenever premiums come close to that level – a legislative change is made, reducing benefits. Within a regime there is a slow upward movement to average premiums. This is to be expected because many benefits are inflation linked (particularly loss of income).

Graph 14 : NSW CTP Premiums



Unfortunately the history of the MAA data is too short for a meaningful statistical analysis of the regime switches, but one would expect that a regime switching model would explain the process fairly well.

Conclusion

Gibbs sampling methodology enables the complicated process of estimating parameters for a regime switching model to be broken down into simpler, more tractable components. The computer time required to estimate the model parameters appears to be much faster than that required for non-linear regressions, as the negative log likelihood function or least squares function to be minimised has many local minima. A further advantage is that as a Bayesian method, it can explicitly allow for information that comes from the professional judgement of the actuary via the use of prior distributions. Finally, gibbs sampling calculates sample values as part of its methodology. These can be used as sample forecast scenarios for pricing as risk management purposes, for reinsurance pricing, and for dynamic financial analysis.

The use of regime switching models enables quite intuitive models to be applied to data. In the case studies presented in this paper, the model is a saw-tooth, a series of straight lines joining a series of peaks and troughs. Non-technical insurance people can easily understand the basis of such models. The results can be presented as probabilities, which make for simple graphical

presentations. There may even be some theoretical basis for using a saw-tooth model to represent the insurance cycle process.

References and Further Reading

Cummins J.D. et al (1991) Cycles and Crises in Property / Casualty Insurance: Causes and Implications for Public Policy. National Association of Insurance Commissioners, Kansas City, MO

Du Plessis Roger M. (1996) Poor Man's Explanation of Kalman Filtering or How I stopped Worrying and Learned to Love Matrix Inversion. Taygeta Scientific Incorporated

Harris G.R. (1999) Markov Chain Monte Carlo Estimation. *Astin Bulletin* Volume 29, No. 1

Kim C.J. and Nelson C.R. (1999) State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. The MIT Press, Cambridge, Massachusetts

Tanner M.A. (1996) Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. Springer-Verlag, New York

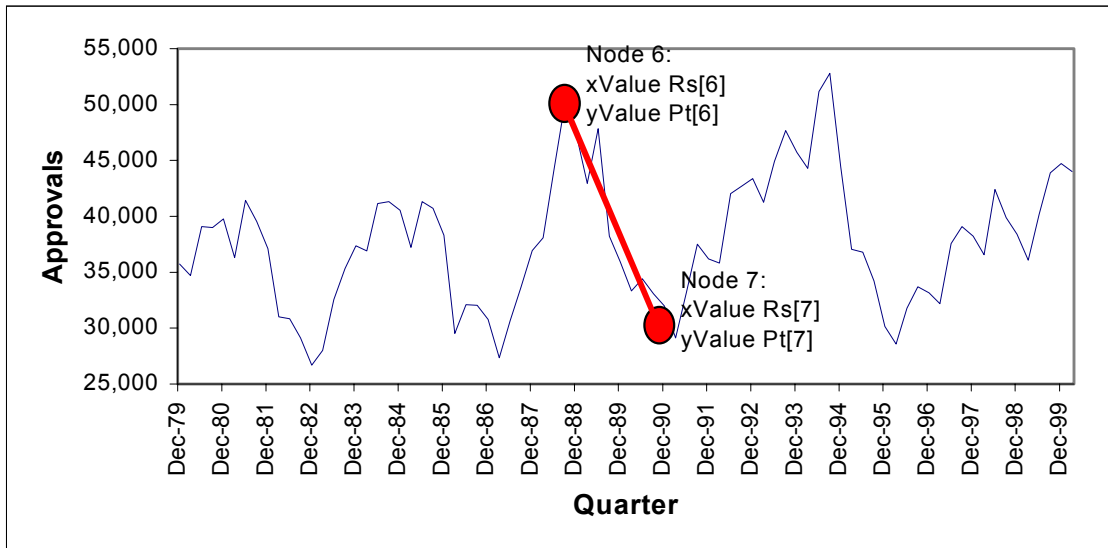
Appendix : Stuff You Will Not Want to Read if You Dislike Algebra and Greek Letters – A Mathematical Description of the Saw-Tooth Regime Switching Model

The Initial Model

The notation for the fitted value at time t is F_t . The notation for the observed value at time t is Y_t .

The initial model will be a series of lines joining the regime switches. Each regime switch will be a node occurring at a particular time and having a particular number of approvals per quarter. The notation for the location of the n^{th} node will be $R_s[n]$ and the notation for the number of approvals per quarter of the n^{th} node will be $P_t[n]$.

Graph 15 : Model Nodes



At points of time between nodes we will linearly interpolate between the nodes, but we will adjust for which quarter of the year we are in, to allow for seasonality. Thus:

$$Y_t = Pt[r] * Lf_t + Pt[r+1] * (1 - Lf_t) + Qf1 * (qtr_t = 1) + Qf2 * (qtr_t = 2) + Qf3 * (qtr_t = 3) + Qf4 * (qtr_t = 4) + \text{Sigma} * e_t$$

Where:

- r is the regime index at time t, beginning at 1
- Lf_t is the linear factor at time t
and $Lf_t = (Rs[r + 1] - t) / (Rs[r + 1] - Rs[r])$
- Qf1 is the seasonal factor for the first quarter
- Qf2 is the seasonal factor for the second quarter
- Qf3 is the seasonal factor for the third quarter
- Qtr_t is the quarter number for time t (e.g. March quarter = 1)
- Sigma is the standard deviation of the error term
- e_t is a random variable $\sim N(0, 1)$

The first data point lies somewhere between a peak and a trough. Rather than extrapolate backwards to the previous node, I have set $Rs[1]$ to equal the first time period and to be a starting node rather than a regime switching node.

Sampling Pt and Qf, given Rs and Sigma and Y

If we know when each node occurs, then the remaining parameters can be estimated via linear regression using maximum likelihood. The dependent value is F_t , the independent variables are the Pt and Qf parameters, and the data is Lf_t and qtr_t .

The notation for the Pt and Qf values is Beta.

The prior distribution is $\text{Beta} \mid \text{Sigma} \sim N(\text{Beta}_0, \text{Sigma}_0)$. We can choose informative priors or uninformative priors. For the purpose of this exercise we are using uninformative priors – so Sigma_0 will contain high covariances to signify that any value is just as likely.

Because of the assumption of normality, the likelihood function of the parameters is given by:

$$L(\text{Beta} \mid \text{Sigma}, Y) = (2 * \pi * \text{Sigma}^2)^{-N/2} * \exp(-1 / (2 * \text{Sigma}^2) * (Y - X * \text{Beta})^T (Y - X * \text{Beta}))$$

where N is the number of data points.

The posterior distribution is $\text{Beta} \mid \text{Sigma}, Y \sim N(\text{Beta}_1, \text{Sigma}_1)$. Where:

$$\text{Beta}_1 = (\text{Sigma}_0^{-1} + \text{Sigma}^{-2} * X^T X)^{-1} (\text{Sigma}_0^{-1} * \text{Beta}_0 + \text{Sigma}^{-2} * X^T Y)$$
$$\text{Sigma}_1 = (\text{Sigma}_0^{-1} + \text{Sigma}^{-2} * X^T X)^{-1}$$

This follows since:

$$f(\text{Beta} \mid \text{Sigma}, Y) \propto f(\text{Beta} \mid \text{Sigma}) * L(\text{Beta} \mid \text{Sigma}, Y)$$

For a full derivation see Kim and Nelson (1999).

While the prior distribution has been uninformative, we can add prior knowledge in other ways. In this paper rejection sampling has been used. Parameter samples that have trough values higher than adjacent peak values have been rejected and a replacement sample taken until an appropriate sample set of parameters is obtained.

Since Sigma_1 is a covariance matrix we should generate sample values for the parameters such that:

$$\text{BetaSample} = \text{Beta}_1 + A * E$$

Where:

$A^T A = \text{Sigma}_1$ (i.e. A is the Cholesky decomposition of Sigma_1)
 E is a vector containing samples from $N(0, 1)$

Sampling Sigma, given Rs and Y and Pt and Qf

Once we know the regime locations and the parameters, we can make an inference on the standard deviation of the error term because we have both Y_t and F_t .

Given that $e_t \sim N(0, 1)$, then
 $\text{Sigma} * e_t \sim N(0, \text{Sigma})$

So $\sum e_t \sim \chi^2(N)$
 and $W = \sum \text{Sigma} * e_t \sim \chi^2(N / 2, 2 / \text{Sigma})$

In order to take advantage of natural conjugate prior distributions, we use a prior:

$$1 / \text{Sigma}^2 | B \sim \Gamma(v_0 / 2, 2 / d_0)$$

The uninformative prior is a small d_0 .

Because of the assumption of normality in the model, the likelihood function is given by:

$$L(1 / \text{Sigma}^2 | \text{Beta}, Y) = (2 * \pi * \text{Sigma}^2)^{-N/2} * \exp(-1 / (2 * \text{Sigma}^2) * (Y - X * \text{Beta})^T (Y - X * \text{Beta}))$$

We multiply the prior density by the likelihood function to get the posterior density.

$$1 / \text{Sigma}^2 | B, Y \sim \Gamma(v_1 / 2, 2 / d_1)$$

where:

$$v_1 = v_0 + N$$

$$d_1 = d_0 + (Y - X * \text{Beta})^T (Y - X * \text{Beta})$$

For a full derivation see Kim and Nelson (1999). Note that they use a different parameterisation of the Gamma to that used in this paper.

Once the gamma distribution parameters have been calculated, the new value for Sigma is generated by sampling from the gamma distribution and taking that sample to the power of minus two.

Sampling $R_s[r]$ (for each r , where $r > 1$ and $r < 12$), given Σ and Y and P_t and Q_f and $R_s[j]$ (for all j , where $j <> r$)

The end nodes are special cases requiring special treatment because there is only a node on one side and because of the possibility that the node may lie outside the data area. However we can easily sample each of the other node locations in turn when we have a node on each side of the sampled node and when all of the node possibilities lie within the data range. Note that we have used an implicit prior – the number of nodes is equal to 12.

We can start by an intuitive test for rejection sampling:

$$R_s[r-1] < R_s[r] < R_s[r+1]$$

We limit the possible choices of node location to those locations that lie between the location of the previous node and the location of the subsequent node.

Furthermore we are limiting the range of the eleventh node (the second last) to lie within the period for which data is available.

For each possible value of $R_s[r] = i$ (there are a finite number of possibilities since $R_s[r]$ can only take integer values) we calculate the probability of $R_s[r]$ taking on that value, given all of the other information.

Using Bayes Theorem:

$$\Pr(R_s[r] = i \mid Y) = L(Y \mid R_s[r] = i) * \Pr(R_s[r] = i) / L(Y)$$

Now we can calculate:

$$L(Y) = \sum_j L(Y \mid R_s[r] = j) * \Pr(R_s[r] = j)$$

$\Pr(R_s[r] = i)$ is a prior distribution. We shall use an uninformative prior (although we will subsequently use some rejection sampling). So this means that the probability that $R_s[r] = i$, given Y is equal to the ratio of the likelihood for location i to the sum of the likelihoods for all possible locations.

$$\Pr(R_s[r] = i \mid Y) = L(Y \mid R_s[r] = i) / \sum_j L(Y \mid R_s[r] = j)$$

As suggested in the previous paragraph further rejection sampling is required. $R_s[r]$ should only take on values for which a peak is higher than the troughs either side, and a trough is lower than the peaks either side. In this way we specify that the model must have a saw-tooth shape rather than a being more generic stepwise linear function.

Once we have the probability of each location value it is a simple matter to generate a uniform random variate and compare that to the cumulative probability function to generate a random location.

Sampling $Rs[12]$, given Σ and Y and Pt and Qf and $Rs[j]$ (for all j , where $j < > 12$)

The twelfth node is special. It lies either on the last data point, or afterwards. The key question is about what can be inferred from the data and the other parameters that would affect the probability of different node locations.

There are three data about the other nodes that could infer something useful:

1. The gaps between the nodes
2. The height of the nodes
3. The type of the last node (peak or trough)

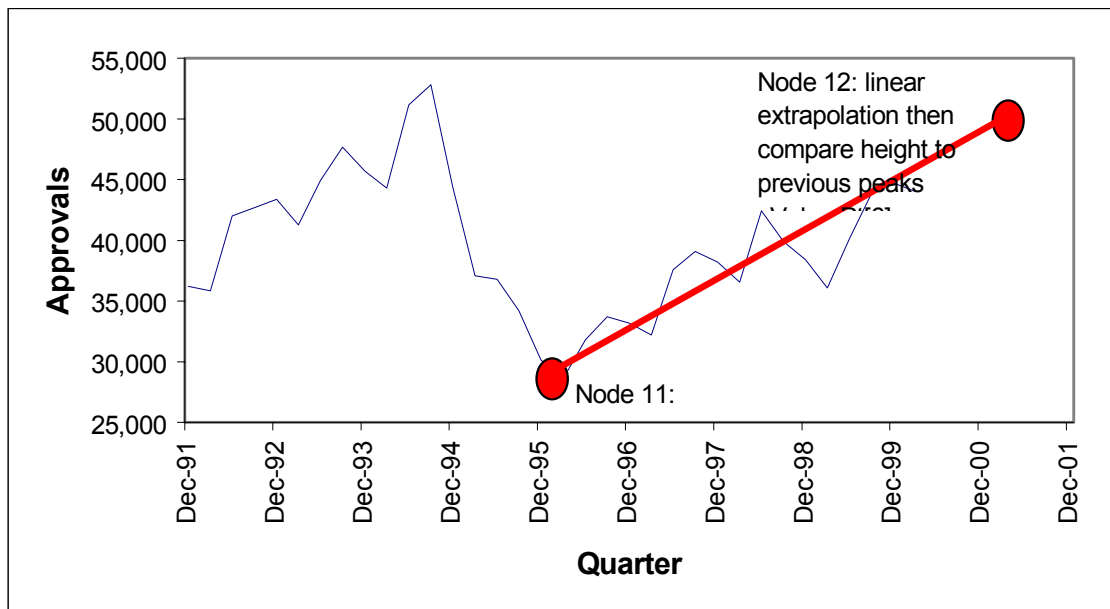
There is one aspect about the data that could infer something useful:

1. The slope of the current trend

Using the slope of the current trend, we can infer the height of the node given its location. We can then compare that height to the distribution of the other height samples to infer the likelihood of that height. We can also compare the gap in location between the twelfth node and the eleventh node to the historical gaps. A negative binomial distribution can be used for this purpose. Thus the likelihood of the last regime lasting a particular time is proportional to the likelihood of the node height for the last node extrapolated from the fitted slope times the probability of the gap between the ninth node and the tenth node.

$$L(Rs[12] = i | Y, Rs[r]) \propto L(\text{extrapolated}(Pt[12]) | Y, Pt[r]) \\ * Pr(Rs[12] - Rs[11] | Rs[r])$$

Graph 16 : Extrapolating a Height for the Twelfth Node



A Second, More Parsimonious Model

While all of the parameters in the initial model were significant, not all were significantly different to one another. In order to make the model more parsimonious, we should combined some of these variables:

- combine the first two trough heights
- combine all subsequent trough heights
- combine the first two peak heights
- combine all subsequent peak heights
- combine the June and September seasonal factors

These changes serve a secondary purpose, in that they simplify the extension of the model to forecast the eleventh and subsequent regime switching points.

Finally, we will use only the length of the growth regimes (rather than both growth and decline regimes) to infer the location of the twelfth node, since it ends a growth phase.

Sampling $R_s[13+]$, given Sigma and Y and P_t and Q_f and R_s

Sampling the location of the thirteenth and subsequent nodes is simpler than sampling the twelfth node because there is no data between the nodes.

The location of the node is taken by sampling from the distribution of the length of the upward or downward regimes and adding that period to the time of the previous node.

The height of the node is the same as the previous trough or peak.