Regime Switching and Insurance Cycles

By Colin Priest BEc FIAA

Synopsis

Traditional time series techniques fail when the duration of historical cycles varies. A great deal of the variability experienced by insurers within the insurance cycle is due to the varying cycle periods.

This paper introduces the use of regime switching models as a means of modelling insurance cycles and applies a saw-tooth model to two sets of historical data with good results.

The process of estimating parameters for a regime switching model can be quite difficult using traditional techniques. A more tractable Gibbs sampling algorithm has been applied to break the model into simple components.

Regime Switching and Insurance Cycles

A Real Life Problem

A client came to me with a problem: a long tailed financial insurance class had occasional bad years, and the client wanted to know whether these bad years were part of a cycle, and if so, could that cycle be predicted so that losses could be recognised earlier (triangle reserving techniques took years to recognise the bad years) and perhaps even prevented or reduced in severity?

It quickly appeared that these claim frequencies were related to the building cycle – the number of building approvals, which fluctuated over time. I then needed to forecast the building cycle in order to forecast the occurrence of bad claim years.

Graph 1: The Building Cycle

The Building Cycle
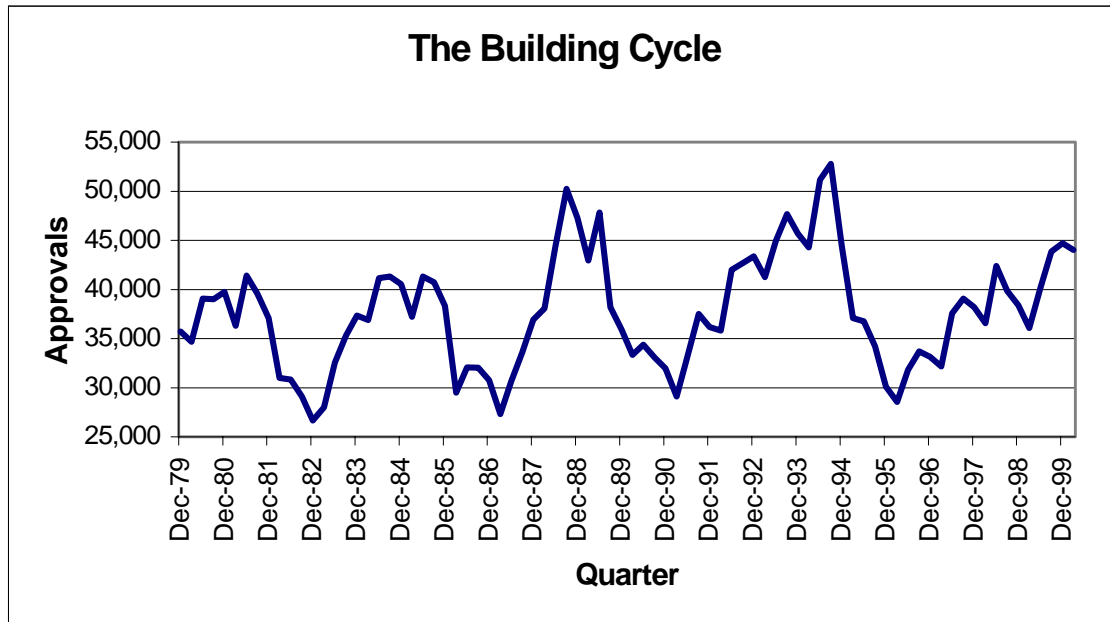
We see in Graph 1 that the pattern looks to be regular with a sort of saw-tooth pattern, but also with some possible seasonal effect within the year.

The Failure of Traditional Forecasting Techniques

The first step was to run the data through some forecasting software. The particular software used 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. A large negative was that I had to tell the software the length of the cycle period in order for it to fit a model.

Graph 2: Traditional Time Series Analysis
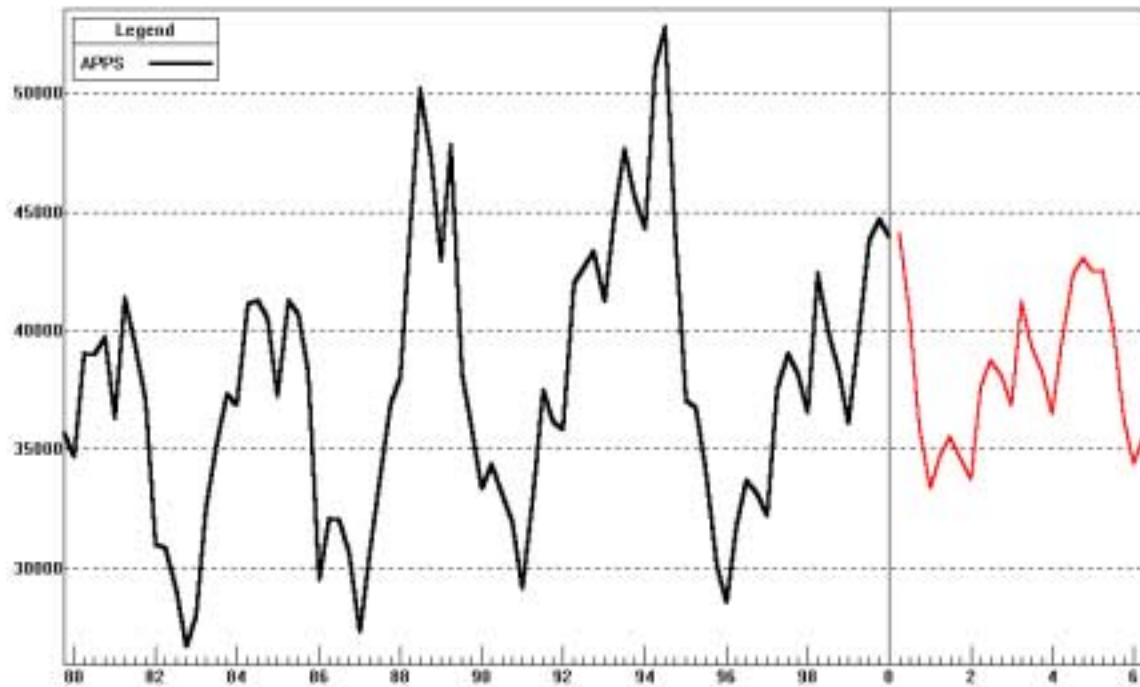
The values in Graph 2 show the model fitted by the software. The output is shown in Table 1. Having to tell the software the length of the cycle is the reverse of what was wanted – we wanted the forecasting method to tell us what the data infers about the cycle length. This approach also has the drawback that each cycle is fixed to exactly the same length. The historical data is not consistent with this assumption, as the past two cycles have been clearly longer than the previous three. Any attempt to price multi-year smoothing reinsurance using this model would underestimate the variability of occurrence of bad years and perhaps estimate an incorrect price as a result. A good forecast model would provide an estimate the variability of cycle lengths.

A further problem is that the forecast does not appear to be obtaining the correct amplitude of the cycles. The troughs of the historical cycles have been at values relatively close to one another (approximately 29,000 approvals per quarter), yet the forecast value of the next trough lies considerably higher than these historical values.

Finally, the historical cycles have different shapes because the location of the peak lies at different durations within the cycle. Yet the traditional model repeats the shape of the last cycle, a slow upward movement followed by a fast downward movement.

Table 1: Output from Traditional Time Series Forecasting

Expert data exploration of dependent variable Apps

Length 82 Minimum 26677.000 Maximum 52803.000
Mean 37713.412 Standard deviation 5739.178

Classical decomposition (multiplicative)
   Trend-cycle: 7.86% Seasonal: 42.39% Irregular: 49.76%

Choice is narrowed down to Box-Jenkins or exponential smoothing.

Box-Jenkins outperforms exponential smoothing by 3004.541 to 5308.574 out-of-sample
Mean Absolute Deviation. I tried 210 forecasts up to a maximum horizon 20.

Series is stationary but seasonal.

Recommended model: Box-Jenkins

Forecast Model for Apps
ARIMA(2,0,0)*(1,0,0)

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	1.1406	0.1064	10.7160	1.0000
a[2]	-0.3507	0.1068	-3.2850	0.9985
A[20]	0.7594	0.0781	9.7296	1.0000
_CONST	1906.6126			

Within-Sample Statistics

Sample size 82   Number of parameters 3
Mean 3.771e+004   Standard deviation 5774
R-square 0.791   Adjusted R-square 0.7857
Durbin-Watson 2.075   * Ljung-Box(18)=33.02 P=0.9834
Forecast error 2673   BIC 2844
MAPE 0.04778   RMSE 2624
MAD 1852

Regime Switching Models

Regime switching models are characterised by a number of 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.

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 cycles between these two regimes. A regime switching model allows the econometrician to measure the different characteristics of growth and recession.

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 regimes for stability and uncertainty.

The reason that regime switching models have only come into use in relatively recent times is that they are different to fit parameters to 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.

Estimation of parameters for regime switching models has become computationally feasible due to the development of Monte Carlo Markov Chain (MCMC) methods, to which the Gibbs Sampler belongs.

Gibbs Sampling

Gibbs-sampling is an MCMC algorithm for approximating joint distributions by sampling from conditional distributions. Suppose you wish to estimate a multivariate probability function $f(x_1, x_2, \ldots, x_N)$ 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 $(z_1, z_2, \ldots, z_N)$
2. Draw a sample from $f(z_1 \mid z_2, \ldots, z_N)$
3. Draw a sample from $f(z_2 \mid z_1, z_3, \ldots, z_N)$
4. Iterate through j (where steps 2 & 3 are the first two iterations) to sample from $f(z_j \mid z_1, z_n, \ldots z_{j-1}, z_{j+1}, \ldots, z_N)$
5. The last iteration of j is $f(z_N \mid z_1, z_2, \ldots, z_{N-1})$

Steps 2 to 5 can be repeated K times. The distribution of the sample values will converge to $f(x_1, x_2, \ldots, x_N)$.

You can store the sample values and use them to approximate $f(x_1, x_2, \ldots, x_N)$. don't store the early sample values because they won't have converged towards the target distribution because the starting values were arbitrary.

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 nth node will be Rs[n] and the notation for the number of approvals per quarter of the nth node will be Pt[n].

Graph 3: Model Nodes
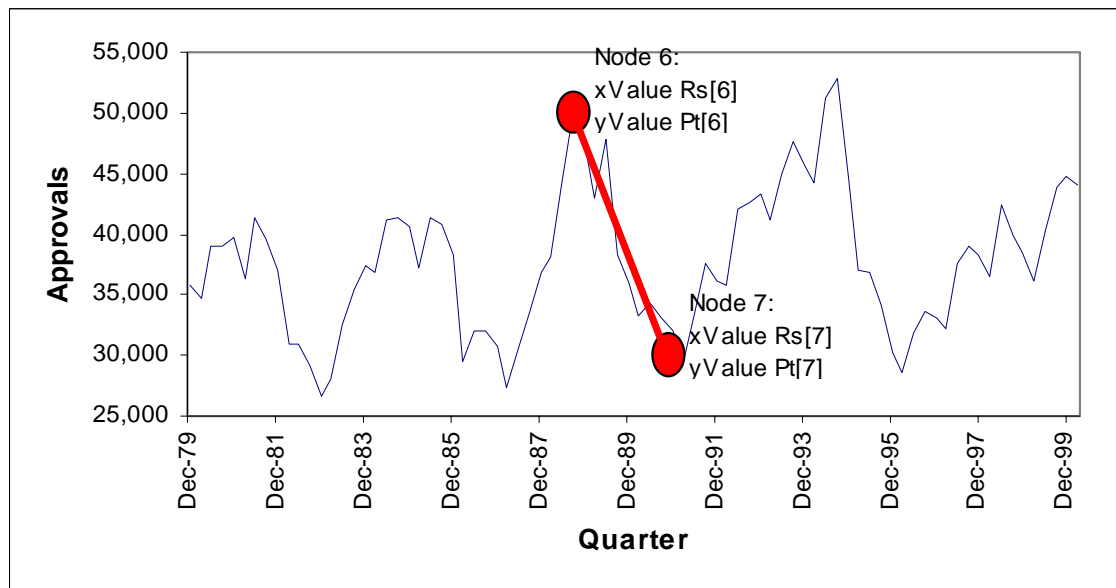
At points of time between nodes we will linearly interpolate between the nodes, but we will adjust for which quarter of the year we are in, to allow for seasonality. Thus:

$$Y_t = Pt[r] * Lf_t + Pt[r+1] * (1 - Lf_t) + Qf1 * (qtr_t = 1) + Qf2 * (qtr_t = 2)$$
$$+ Qf2 * (qtr_t = 2) + Sigma * e_t$$

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 ~ 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 Beta | Sigma $\sim$ N(Beta$_0$, 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(Beta | Sigma, Y)
$$= (2 * \pi * Sigma^2)^{-N/2} * \exp(-1 / (2 * Sigma^2) * (Y - X * Beta)^T (Y - X * Beta)$$

where N is the number of data points.

The posterior distribution is Beta | Sigma, Y $\sim$ N(Beta$_1$, Sigma$_1$). Where:

$$Beta_1 = (Sigma_0^{-1} + Sigma^{-2} * X^T X)^{-1} (Sigma_0^{-1} * Beta_0 + Sigma^{-2} * X^T Y)$$
$$Sigma_1 = (Sigma_0^{-1} + Sigma^{-2} * X^T X)^{-1}$$

This follows since:

f(Beta | Sigma, Y) α f(Beta | Sigma) * L(Beta | 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:

BetaSample = Beta$_1$ + A * E

Where:

$A^T A = Sigma_1$ (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
Sigma * $e_t \sim N(0, Sigma)$

So $\Sigma e_t \sim \chi^2(N)$
and $W = \Sigma$ Sigma * e_t) $\sim \Gamma(N / 2, 2 / Sigma)$

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

$1 / Sigma^2 \mid B \sim \Gamma(v_0 / 2, 2 / d_0)$

The uninformative prior is a small d_0.

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

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

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

$1 / Sigma^2 \mid B, Y \sim \Gamma(v_1 / 2, 2 / d_1)$

where:
$v_1 = v_0 + N$
$d_1 = d_0 + (Y - X * Beta)^T (Y - X * Beta)$

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 Rs[r] (for each r, where r > 1 and r < 10), given Sigma and Y and Pt and Qf and Rs[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 10.

We can start by an intuitive test for rejection sampling:

Rs[r-1] < Rs[r] < Rs[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 ninth node (the second last) to lie within the period for which data is available.

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

Using Bayes Theorem:

Pr(Rs[r] = i | Y) = L(Y | Rs[r] = i) * Pr(Rs[r] = i) / L(Y)

Now we can calculate:
L(Y) = Σ_j L(Y | Rs[r] = j) * Pr(Rs[r] = j)

Pr(Rs[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 Rs[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(Rs[r] = i | Y) = L(Y | Rs[r] = i) / Σ_j L(Y | Rs[r] = j)

As suggested in the previous paragraph further rejection sampling is required. Rs[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[10], given Sigma and Y and Pt and Qf and Rs[j] (for all j, where j < > 10)

The tenth 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 tenth node and the ninth 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.

$$L(\text{Rs}[10] = i \mid Y, \text{Rs}[r]) \propto L (\text{extrapolated}(\text{Pt}[10]) \mid Y, \text{Pt}[r])$$
$$* \Pr (\text{Rs}[10] - \text{Rs}[9] \mid \text{Rs}[r])$$

Graph 4: Extrapolating A Height For the Tenth Node
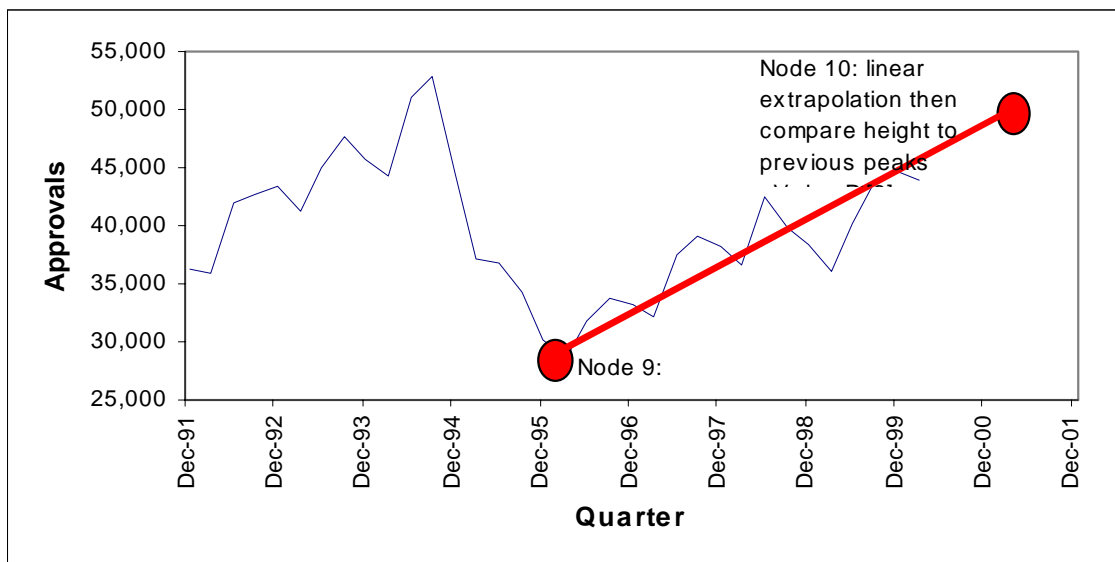
Initial Model Results

Table 2: Initial Model Results

Variable	Mean	Std Dev	Median	Mode
Rs[2]	5.120	0.675	5	5
Rs[3]	10.407	0.509	10	10
Rs[4]	18.098	0.746	18	18
Rs[5]	28.865	0.374	29	29
Rs[6]	34.138	0.345	34	34
Rs[7]	42.064	0.398	42	42
Rs[8]	57.947	0.224	58	58
Rs[9]	62.223	0.421	62	62
Rs[10]	81.903	1.943	81	80
Sigma	1,620	160	1,606	
Qf1	-2,642	368	-2,638	
Qf2	1,123	386	1,128	
Qf3	1,333	391	1,332	
Pt1	36,732	1,121	36,789	
Pt2	40,698	1,029	40,669	
Pt3	26,920	1,067	26,910	
Pt4	42,353	882	42,347	
Pt5	28,202	946	28,203	
Pt6	49,156	936	49,168	
Pt7	30,633	747	30,641	
Pt8	50,620	780	50,624	
Pt9	30,031	801	30,030	
Pt10	45,917	1,765	45,595	
Delta Rs[i]	9.598	5.279	8	5
Upwards	12.131	5.992	13	5
Downwards	7.064	2.645	7	4

The model provides quite robust estimates of the parameters with run times much faster than that required to solve a non-linear least squares regression. Note that the heights of the troughs are quite close to one another.

The seasonal parameters suggest that the December and March quarters have lower building approvals and that the June and September quarters are not significantly different from one another. This makes intuitive sense, because in Australia the summer holiday break (including the 6 week long school holiday break) includes the closure of much of the building industry.

Upward trends have had both a longer average duration and more variability in duration than downward trends. This, combined with the lower variability of trough heights, means that we can forecast downward trends with more certainty than upward trends.

Graph 5: Historical Data vs Average Fitted Values
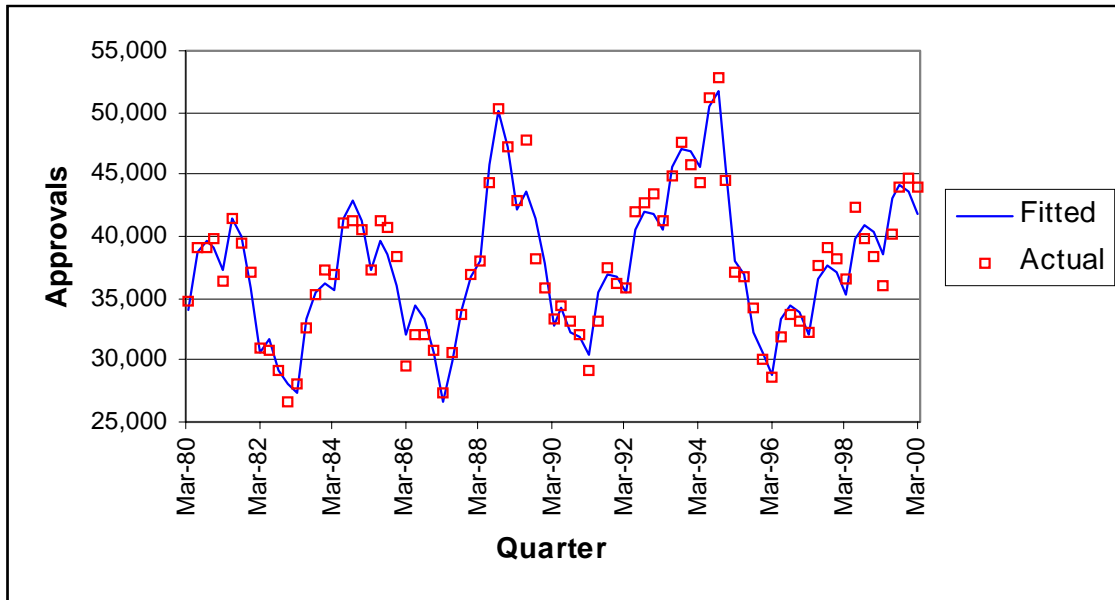
The fitted values for the history show a remarkably good match to the data. A great deal of the apparent variability in the historical data was due to seasonal effects. I admit to being surprised at how well the history is described by a series a straight lines with seasonal adjustment.

Graph 6: Probabilities of Different Locations of the Next Regime Switch
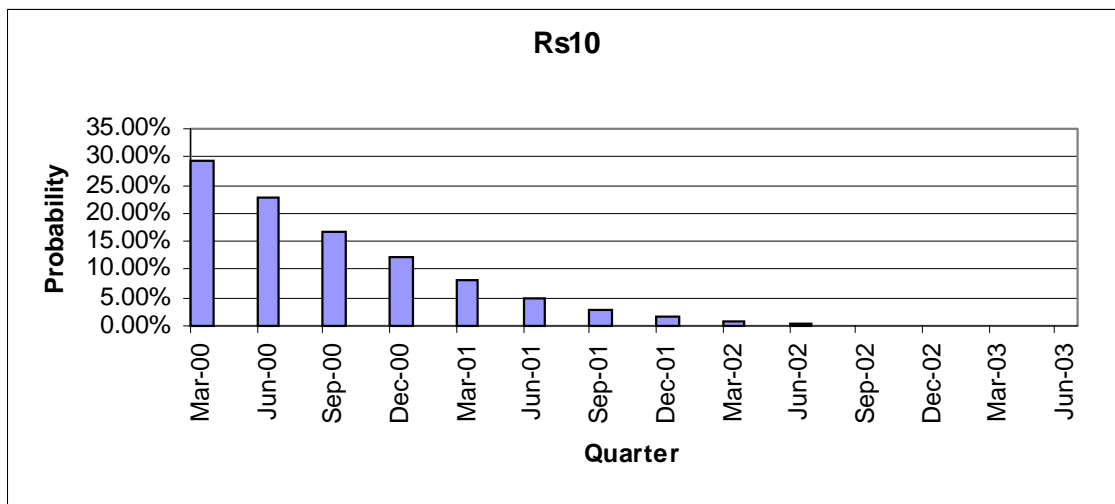
The model predicts that it is almost certain that the current growth trend will finish during the year 2000.

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 are going to include some prior knowledge into the model. From 1st July 2000 a goods and services tax (a type of VAT) will be charged in Australia. This will increase the cost of building construction. People have been moving construction activity earlier to avoid paying the tax. So it is expected that approval levels will drop after June 2000. The monthly approval figures (available for April) also suggest a drop off. So the model will use a prior distribution for Rs[10] as below:

Pr(Rs[10] = March 2000) = 0.5
Pr(Rs[10] = June 2000) = 0.5

This is identical to using rejection sampling for locations of Rs[10] after June 2000.

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

## Sampling Rs[11+], given Sigma and Y and Pt and Qf and Rs

Sampling the location of the eleventh and subsequent nodes is simpler than sampling the tenth 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.

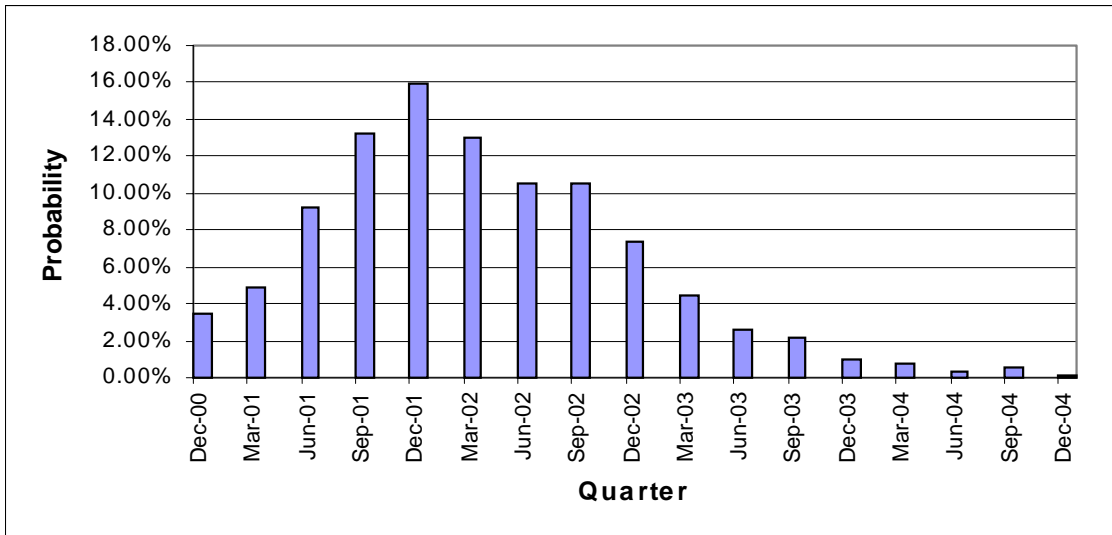The height of the node is the same as the previous trough or peak.

Final Model Results

Graph 7: Final Model
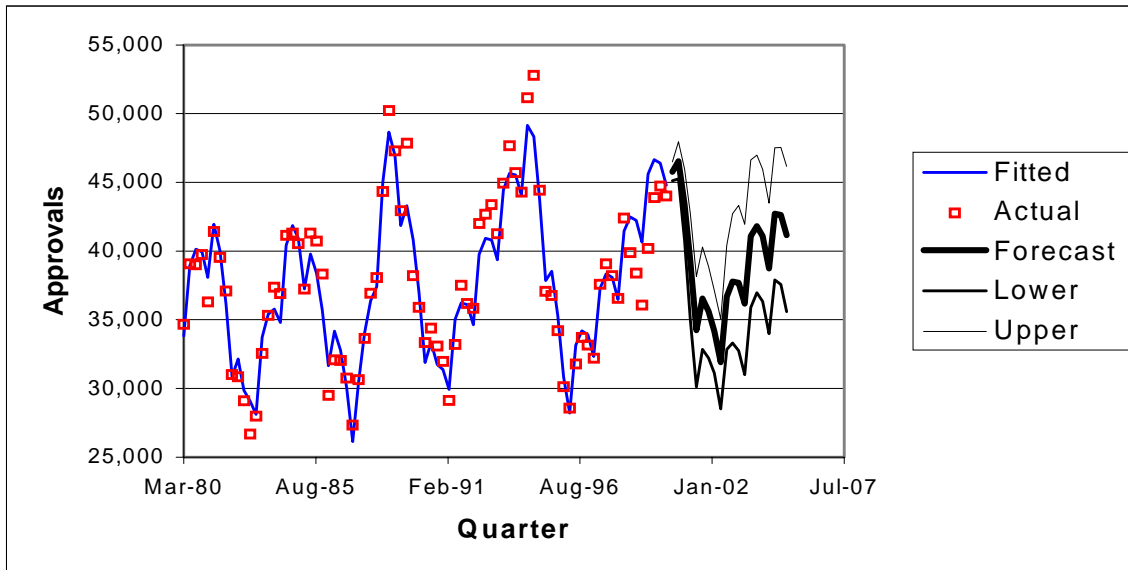
The confidence interval for the forecast broadens rapidly after the next trough in the cycle due to the great variability in the historic growth phases.

Graph 8: Location of the Next Trough

The model not only gives me an estimate of the next trough, but also the probability of different times.

Best of all, the model automatically produces scenario samples from the joint distribution of all of the parameters. These scenarios can be used to price smoothing reinsurance, and as worked examples of the effects of the cover.

Graph 9: Sample Scenarios
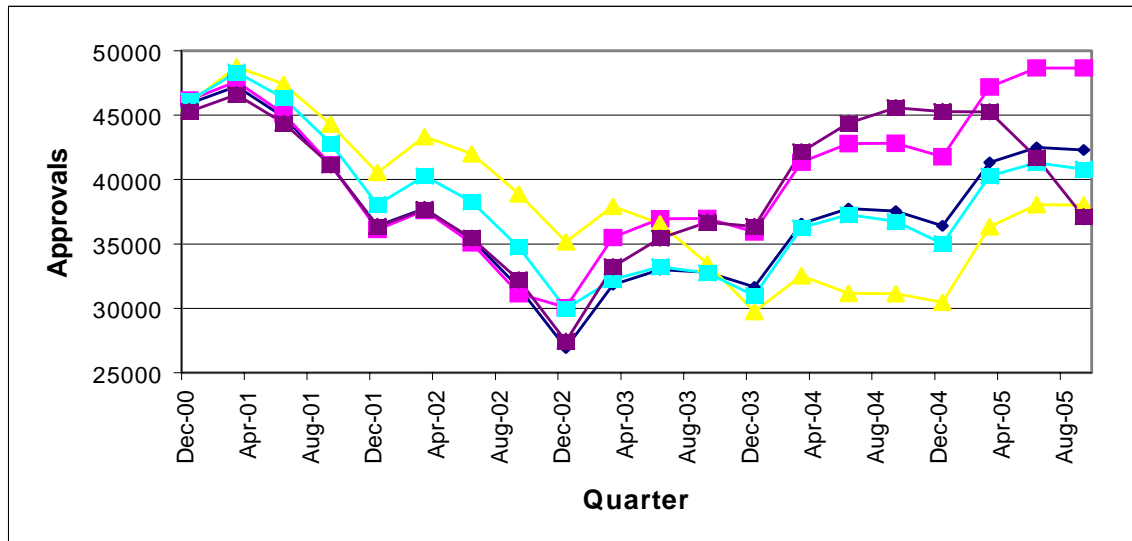
The parameters of the final model are shown in Table 3.

Table 3: Parameter Estimates of Final Model

Variable	Mean	Std Dev	Median	Mode
Rs[2]	4.707	0.928	5	5
Rs[3]	10.229	0.568	10	10
Rs[4]	18.637	0.786	19	19
Rs[5]	28.607	0.540	29	29
Rs[6]	34.362	0.490	34	34
Rs[7]	41.853	0.525	42	42
Rs[8]	57.495	0.524	58	58
Rs[9]	63.263	0.516	63	63
Rs[10]	80.981	0.136	81	81
Rs[11]	88.000	2.850	88	87
Rs[12]	99.870	6.367	99	97
Rs[13]	107.236	7.475	106	108
Sigma	1,994	191	1,978	
Qf1	-2,633	460	-2,641	
Qf2	1,301	341	1,300	
Pt1	36,455	1,491	36,508	
Pt2	41,692	936	41,689	
Pt3	27,892	922	27,887	
Pt4	48,445	682	48,439	
Pt5	30,057	700	30,057	

Application to US Combined Ratios

Cummins et al (1991) present a history of combined ratios in the US, summarised in Graph 10 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 tough to peak and from peak to trough.

Graph 10: US Combined Ratios
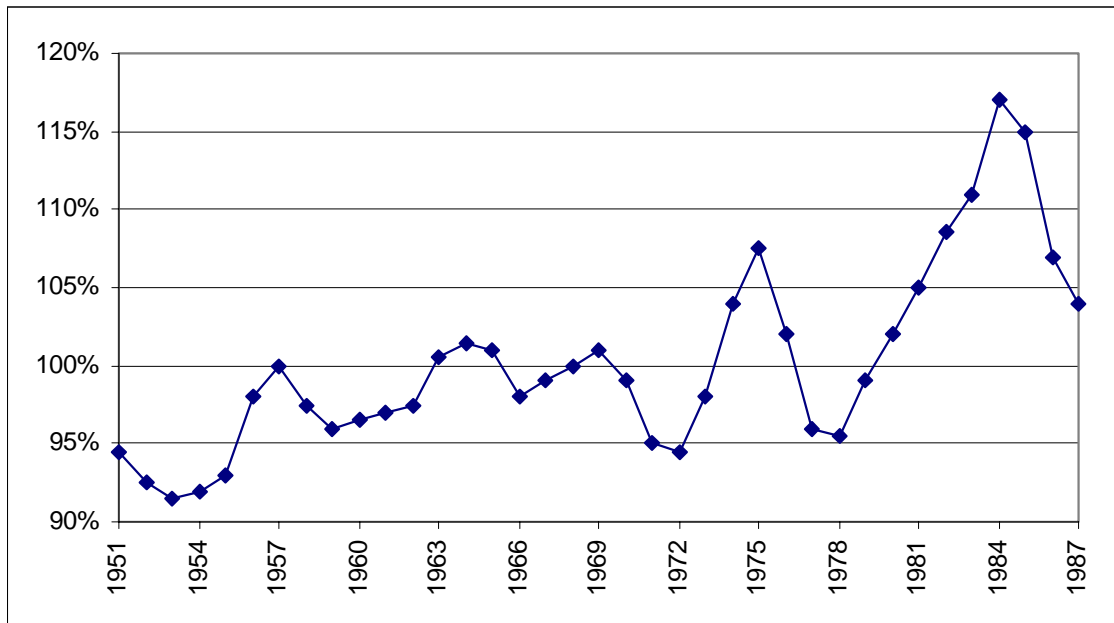
The starting year of data (1951) and the last year of data (1987) both are allocated nodes that are held fixed, much like the starting period for the building cycle example. These two nodes are neither peaks nor troughs. They just exist at the locations at which the data has been censored.

The node locations and heights are then generated in the same manner as for the building cycle data. In this example we are only interested in the location of the nodes, not the height, but the height of the nodes is required in order to generate the location of the nodes.

Table 4: Location of Nodes for US Combined Ratios

Variable	Mean	Std Dev	Median	Mode
Rs[1]	0.000	0.000	0	0
Rs[2]	2.942	0.782	3	3
Rs[3]	5.811	0.482	6	6
Rs[4]	8.811	1.009	9	8
Rs[5]	13.279	0.672	13	13
Rs[6]	15.452	0.649	15	15
Rs[7]	17.790	0.482	18	18
Rs[8]	20.937	0.267	21	21
Rs[9]	23.994	0.079	24	24
Rs[10]	26.727	0.447	27	27
Rs[11]	33.145	0.352	33	33
Rs[12]	36.000	0.000	36	36

Note that the standard deviation of the estimates varies between nodes. 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 11: Fitted Model vs History
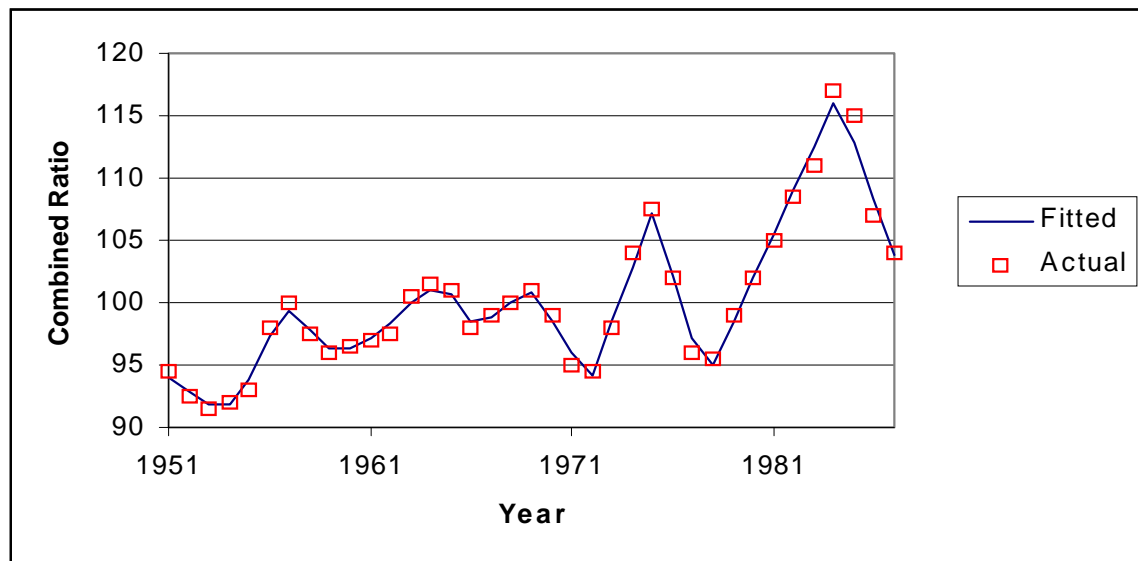
Once again, the saw-tooth model, while simple, provides a good explanation of the behaviour of the cycle. Over short periods of time (two to three years) the saw-tooth model is the most logical because some insurance changes happen in a saw-tooth manner:

- in long tail business the results come in slowly, so the insurer's perception of the risk changes gradually
- premiums earn over a period of time, so even if new information meant that you could accurately estimate the risk price and quality, reported results such as the combined ratio are still affected by past price and quality linearly over the earning period

- due to the inability of measuring the underlying factors, insurers are uncertain of the extent that their results are due to good management or good luck, or bad management or bad luck. It takes a few years of similar or trending results for them to be certain enough to change their minds. This then prompts the regime switch between production goals and quality goals.

Graph 12: Distribution of Periods from Trough to Peak vs Peak To Trough
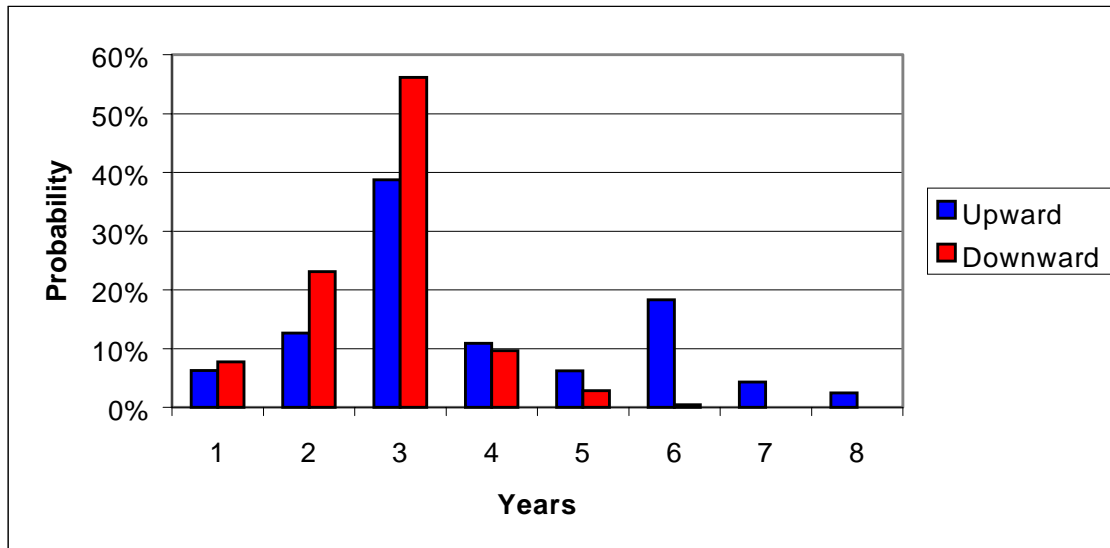
The modal times for both upward trends and downward trends are both a value of three, and the mode is quite dominant, more dominant than one would expect if the underlying distribution was one of the commonly used statistical distributions. It would be a good topic for further research to explain the significance of the three year period in US insurance practice.

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 taken to estimate the model parameters appears to be much faster than that required for non-linear regression, as the likelihood or least squares function to be minimised has many local minima. A further advantage is that it explicitly allows 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 and risk management purposes.

The use of regime switching models enables quite intuitive models to be applied to data. In the case studies presented in this study, 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 model the insurance cycle.

Analysis of historical insurance data shows that the insurance cycle varies in length. Regime switching models are more appropriate for modelling the insurance cycle than traditional time series techniques because they allow the actuary to make inferences about the variability of cycle periods as well as the variability of the heights of the peaks and troughs.

References

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

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