Hidden Markov models: some examples of their application and reflections on their use

Peter Thomson
Statistics Research Associates Ltd

1 Overview

2 What are HMM models?

3 Fitting HMM models

4 Applications:

- Share prices
- GDP growth rates
- Multisite daily rainfall
- Other

5 Concluding remarks

1 Overview

Statistics: the art of applying science

Using selected applications, consider issues such as:

- advantages/limitations of HMMs in practice
- the need to explore/exploit HMM structure
- risk and point forecasting with HMMs
- modelling diverse time/space scales with HMMs
- opportunities for enhanced interpretation and more physically based models

2 What are HMM models?

A hidden Markov model (HMM)

- blocks time series data into consecutive periods of time (regimes)
- has regimes that are not directly observed (they are unobserved, hidden or latent)
- uses simple (dynamic) models within regimes
- models regime switching using Markov chains or variants thereof
- involves at least two (dynamic) time scales; within and across regimes

An HMM provides a simple, readily understood, structure that can be used in its own right or as an exploratory tool to help specify alternative models.

Example

Assume that a stationary macroeconomic or financial time series Y_t follows the model

$$Y_t = \mu_{S_t} + \sigma_{S_t} X_t \quad (t = 0, \pm 1, \ldots)$$

where

- the states S_t form an unobserved stationary Markov chain that takes on the values $1, \ldots, N$

- X_t is a zero mean, unit variance, Gaussian $AR(1)$ process, independent of S_t, that satisfies

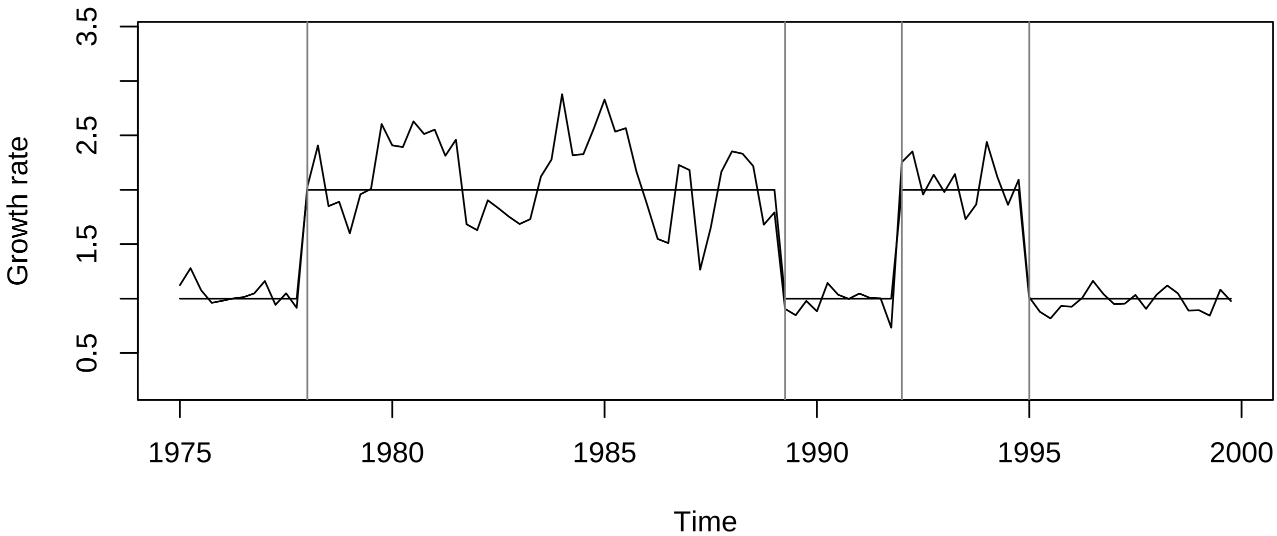
$$X_t = \rho X_{t-1} + \epsilon_t \quad (|\rho| < 1; t = 0, \pm 1, \ldots)$$

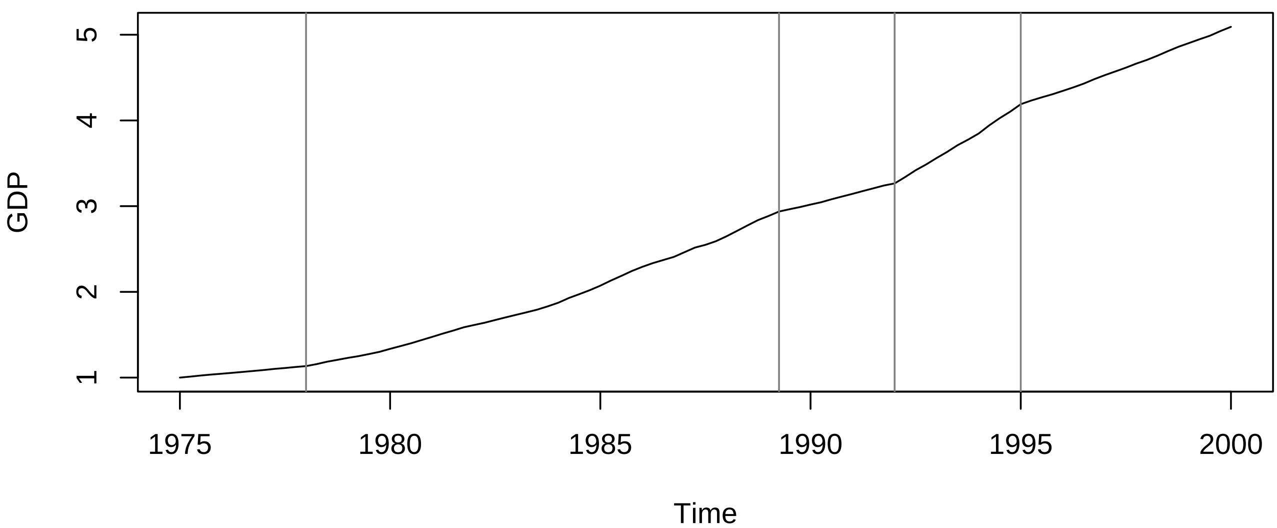
Given S_t, the mean level and volatility

$$E(Y_t|S_t) = \mu_{S_t}, \quad \mathrm{Var}(Y_t|S_t) = \sigma_{S_t}^2$$

switch between the values $(\mu_1, \sigma_1^2), \ldots, (\mu_N, \sigma_N^2)$ according to S_t.

Simulated quarterly GDP growth rates Y_t with $N = 2$ hidden persistent states, $\mu_1 < \mu_2$, $\sigma_1 < \sigma_2$ and $\rho > 0$.





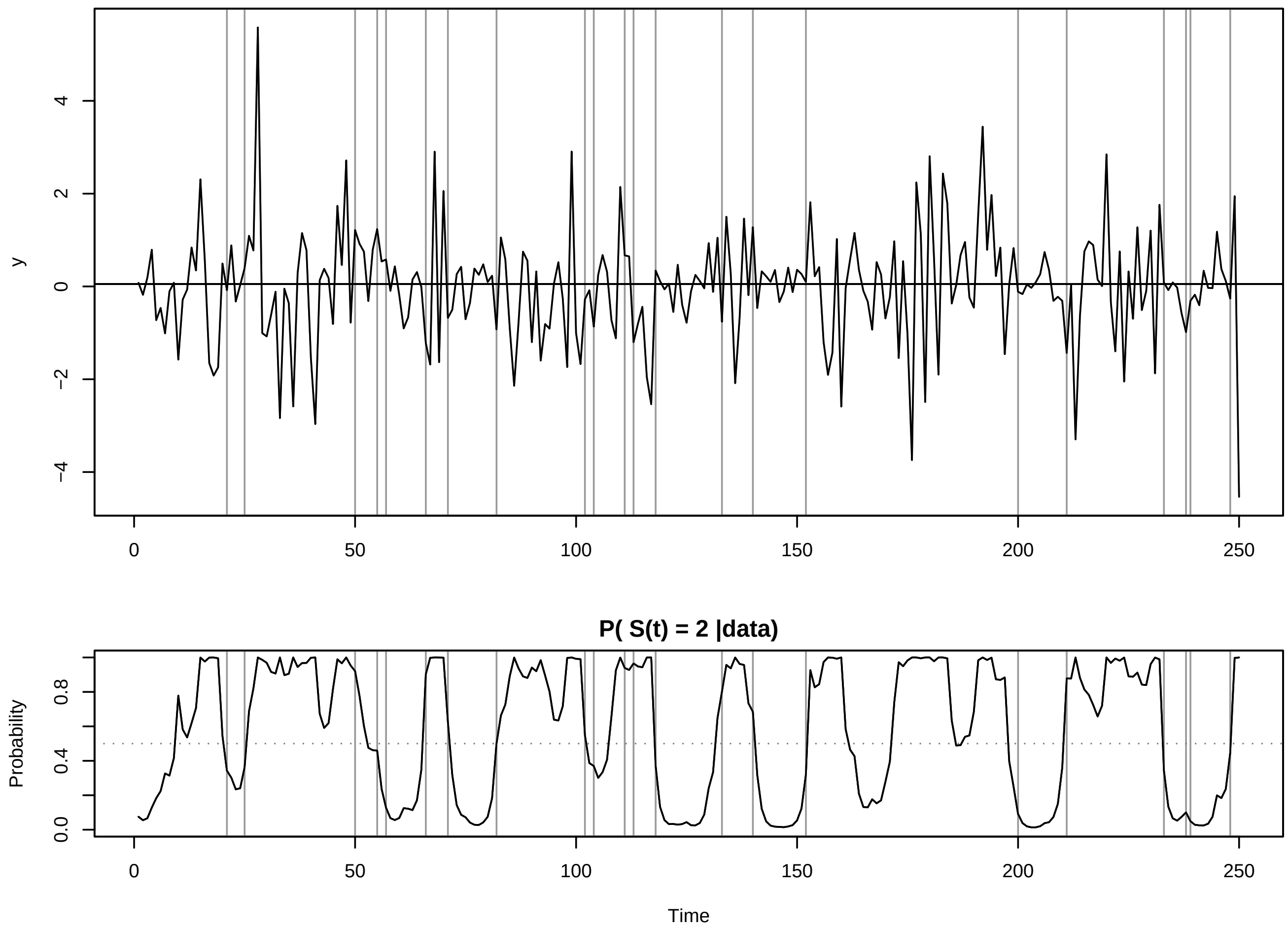
Upper plot shows Y_t with μ_{S_t} (horizontal lines).

Lower plot shows corresponding GDP series.

Vertical grey lines show when S_t changes state.

Regime switching clearly evident!

Simulated daily returns Y_t with $N = 2$ hidden persistent states, $\mu_1 = \mu_2$, $\sigma_1 < \sigma_2$ and $\rho = 0$.



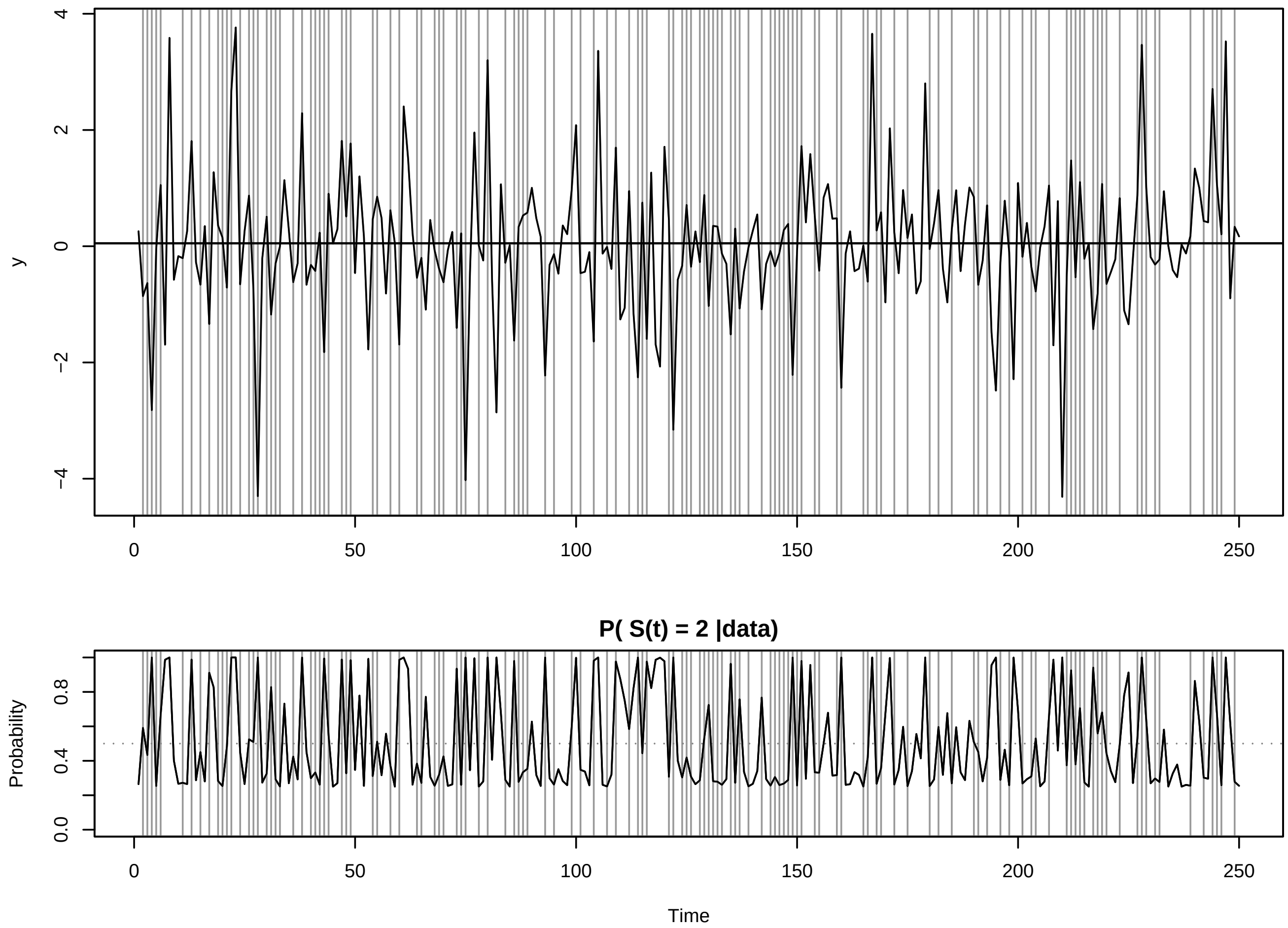
Upper plot shows Y_t with $\mu_1 = \mu_2$ (horizontal line).

Lower plot shows $P(S_t = 2|\mathbf{Y}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes true parameters.

Vertical grey lines show when S_t changes state.

Regime switching evident.

Simulated daily returns Y_t with $N = 2$ hidden independent states, $\mu_1 = \mu_2$, $\sigma_1 < \sigma_2$ and $\rho = 0$.



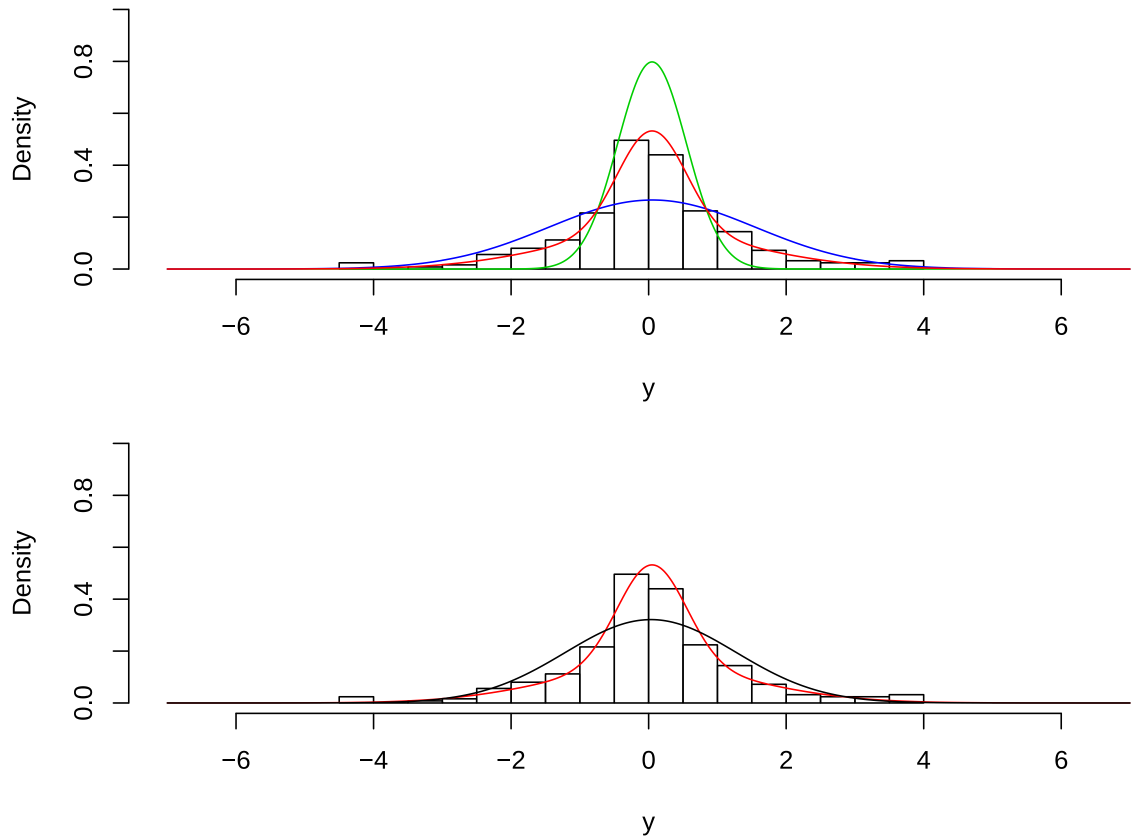
Upper plot shows Y_t with $\mu_1 = \mu_2$ (horizontal line).

Lower plot shows $P(S_t = 2|\mathbf{Y}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes the true parameters.

Vertical grey lines show when S_t changes state.

Regime switching less evident.

Histograms of simulated daily returns Y_t with $N = 2$ hidden independent states, $\mu_1 = \mu_2$, $\sigma_1 < \sigma_2$, $\rho = 0$.



Upper plot shows histogram of Y_t (black) with its theoretical density (red), density when $S_t = 1$ (green), and density when $S_t = 2$ (blue).

Lower plot shows histogram of Y_t (black) with its theoretical density (red), and density of best fitting Gaussian(black).

Same theoretical mixture densities for both independent and dependent examples.

3 Fitting HMM models

HMM is typically fitted to observations $Y_1, \ldots, Y_T$ using maximum likelihood.

Strategy: for previous example and more generally.

- Use EM algorithm to explore log likelihood and obtain a range of suitable initial estimates.
- Starting from initial estimates, use numerical procedures to directly maximise log likelihood.
- Explore structure of HMM by fitting a suitable range of reduced models to the data.
- Examine resulting estimates, AIC values, and a wide range of graphical diagnostics, etc to assess goodness of fit.

Comments:

- Strategy utilises EM algorithm's relative robustness to choice of initial values.

- Likelihood values and EM algorithm depend on quantities such as

$$\gamma_t(j) = P(S_t = j|\mathbf{Y})$$

where $\mathbf{Y}$ denotes the data $Y_1, \ldots, Y_T$.

- The $\gamma_t(j)$ can be calculated efficiently using the forward–backward algorithm of Baum et al (1970). [Code needs to be written with care!]

The $\gamma_t(j)$ are also used to identify likely states and estimate hidden quantities such as

$$E(\mu_{S_t}|\mathbf{Y}) = \sum_{j=1}^{N} \mu_j \gamma_t(j), \quad E(\sigma^2_{S_t}|\mathbf{Y}) = \sum_{j=1}^{N} \sigma^2_j \gamma_t(j)$$

and forecast risk parameters such as

$$P(Y_{T+t} > y|\mathbf{Y})$$

where $T + t$ denotes some future time point.

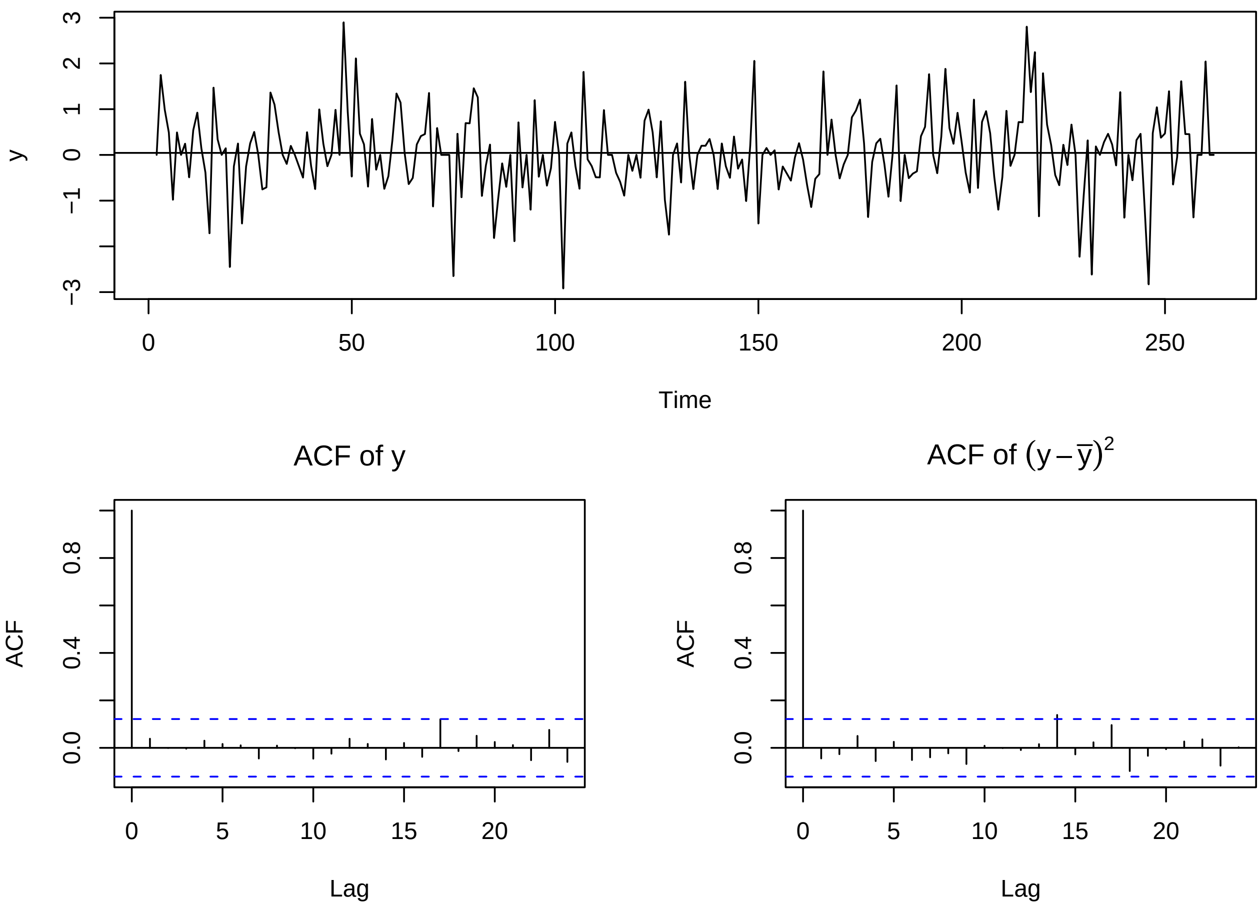
4 Applications

- **Share prices** Consider the daily returns

$$Y_t = \log P_t - \log P_{t-1}$$

of ANZ share prices P_t on NZSE for 2004.

ANZ daily returns for 2004



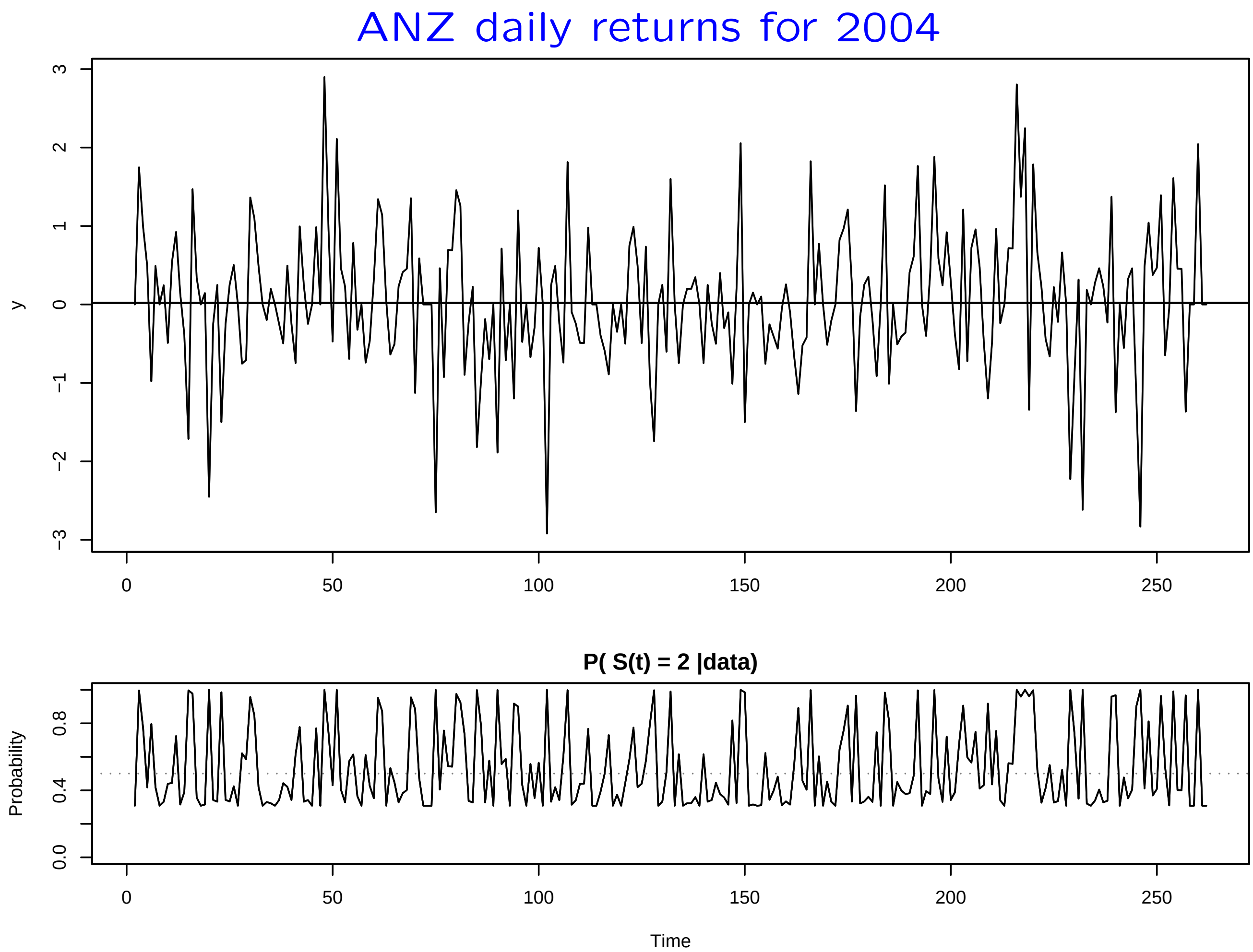
Upper plot shows Y_t with mean (horizontal line).

No evidence of autocorrelation (lower left plot) or volatility clustering (lower right plot).

Fitted HMM is

$$Y_t = \mu + \sigma_{S_t} X_t$$

where X_t are iid $N(0,1)$ and S_t has 2 indep states.

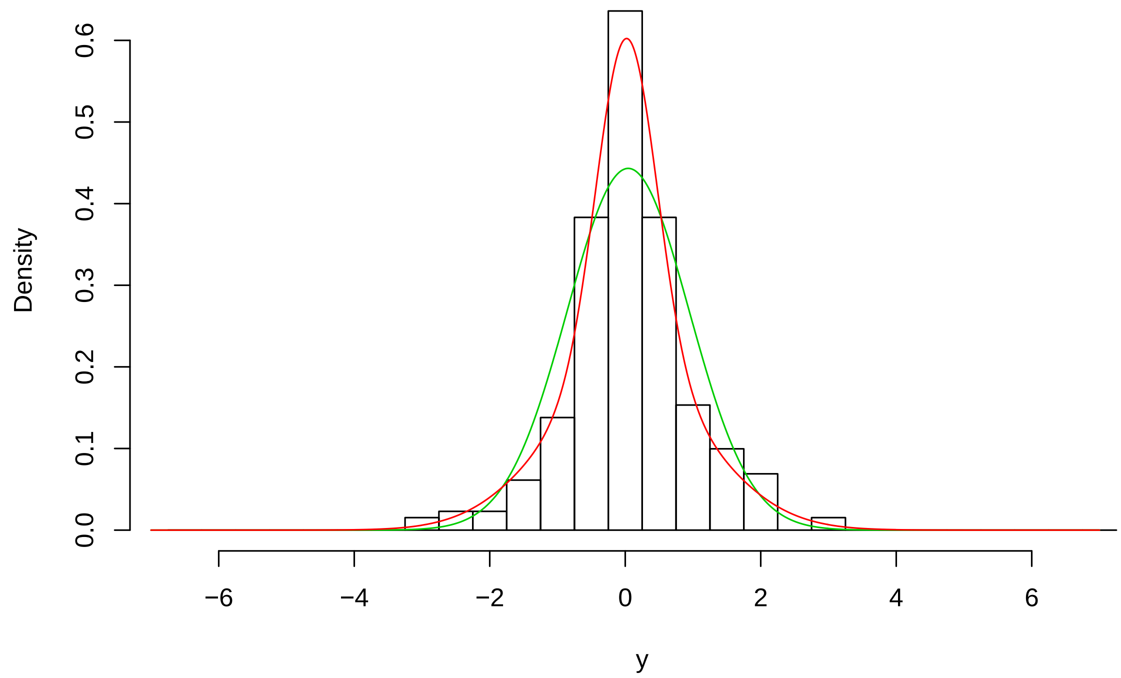


Upper plot shows Y_t with mean (horizontal line).

Lower plot shows $P(S_t = 2|\mathbf{Y}, \widehat{\boldsymbol{\theta}})$ where $\widehat{\boldsymbol{\theta}}$ denotes the estimated parameters.

Independent non–Gaussian returns.

ANZ daily returns for 2004



Histogram of Y_t (black) with fitted HMM density (red) and density of best fitting Gaussian (green).

HMM Gaussian mixture close to a scaled t_4, but HMM has useful structure.

HMM mixture distributions can be used to model many shapes including the leptokurtic, heavy–tailed distributions met in finance.

- **GDP growth rates (Buckle et al, 2002)**

Consider the quarterly growth rates

$$Y_t = \log L_t - \log L_{t-1}$$

of NZ aggregate GDP L_t from 1978–2002.

Fitted HMM is

$$Y_t = \mu_{S_t} + \sigma_{S_t} X_t$$

where X_t is a Gaussian $AR(1)$ with

$$E(X_t) = 0, \qquad \text{Var}(X_t) = 1$$

and X_t is independent of the unobserved, $N = 4$ state, Markov chain S_t.

In principle S_t is specified by the $N(N-1)$ transition probabilities

$$P_{ij} = P(S_{t+1} = j | S_t = i)$$

since $\sum_j P_{ij} = 1$. For $N = 4$ this yields 12 parameters to estimate for S_t.

Too expensive and a major weakness!!

Building on McConnell and Perez–Quiros (2000), S_t modelled as 1–1 function of two independent, 2 state, Markov chains C_t and V_t where

S_t	C_t	V_t	Growth regime	Volatility regime	μ_{S_t}	σ_{S_t}
1	0	0	Low	Low	μ_1	σ_1
2	0	1	Low	High	μ_2	σ_2
3	1	0	High	Low	μ_3	σ_3
4	1	1	High	High	μ_4	σ_4

and $\mu_1 \leq \mu_2$, $\sigma_1 \leq \sigma_2$.

S_t has only 4 free parameters compared to 12 for the general 4 state chain.

Note that

$$S_t = 1 + 2^0 V_t + 2^1 C_t$$

so idea can be extended to $N = 2^p$ state chains resulting in $O(\log N)$ parameters rather than $O(N^2)$.

This structural Markov chain is parsimonious and may provide a suitable explanation of underlying economic or physical process.

Adopt a more general view and

- consider S_t and components C_t, V_t as approximating a general 4 state Markov chain with 12 free parameters [How good is approximation?]

- classify states to regimes within this framework

This allows a more flexible and parsimonious framework within which to explore structure of GDP growth rates.

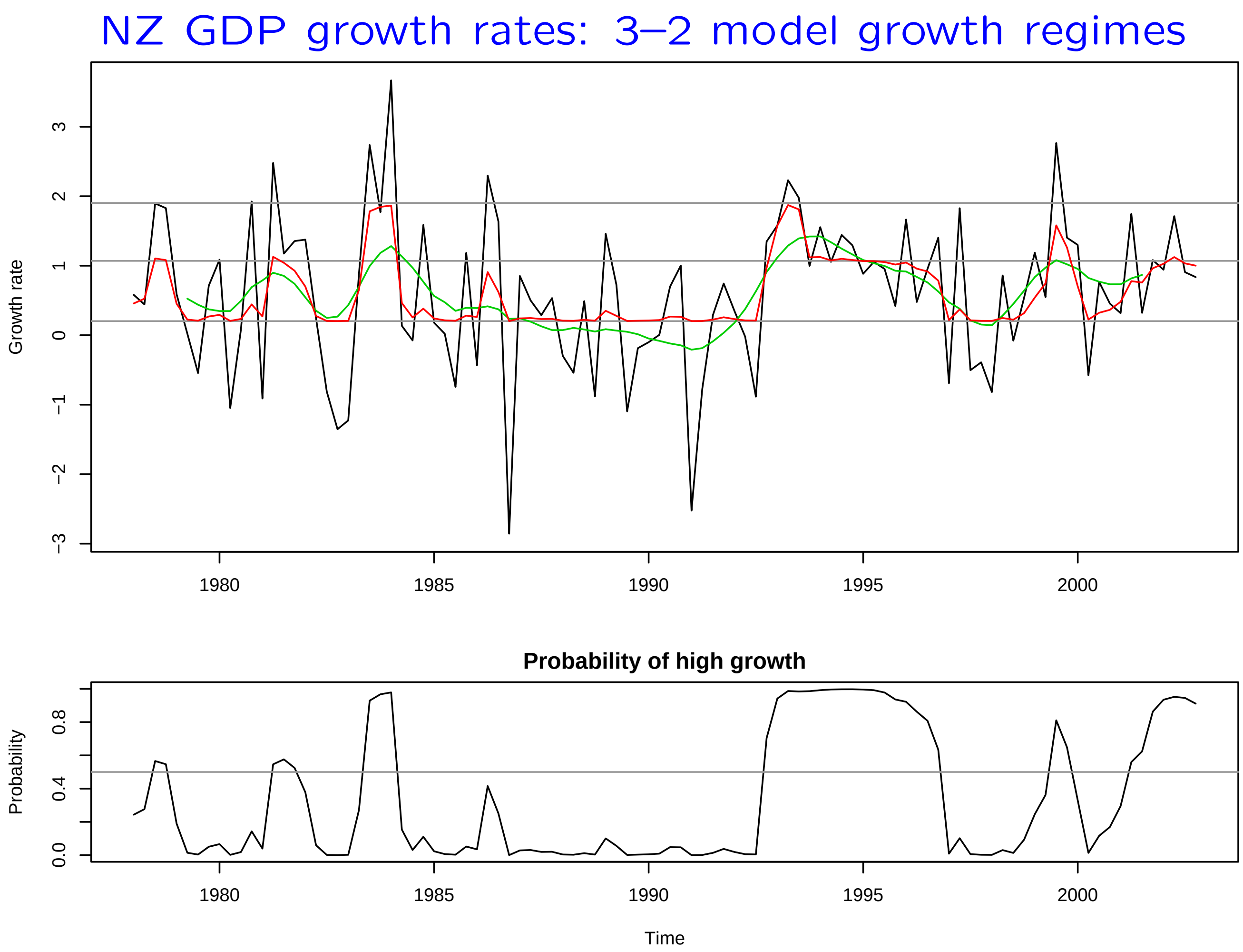
A number of reduced models were fitted and two competing models identified.

- 1 state simple linear $AR(1)$ model with $\mu_1 = \mu_2 = \mu_3 = \mu_4$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$

- 4 state non–linear 3–2 model with $\mu_3 = \mu_4$ (3 means), $\sigma_1 = \sigma_3$, $\sigma_2 = \sigma_4$ (2 variances)

The AIC values for these models have local minima with $AR(1)$ having the absolute minimum. The $AR(1)$ is also a best fitting $ARMA$ model.

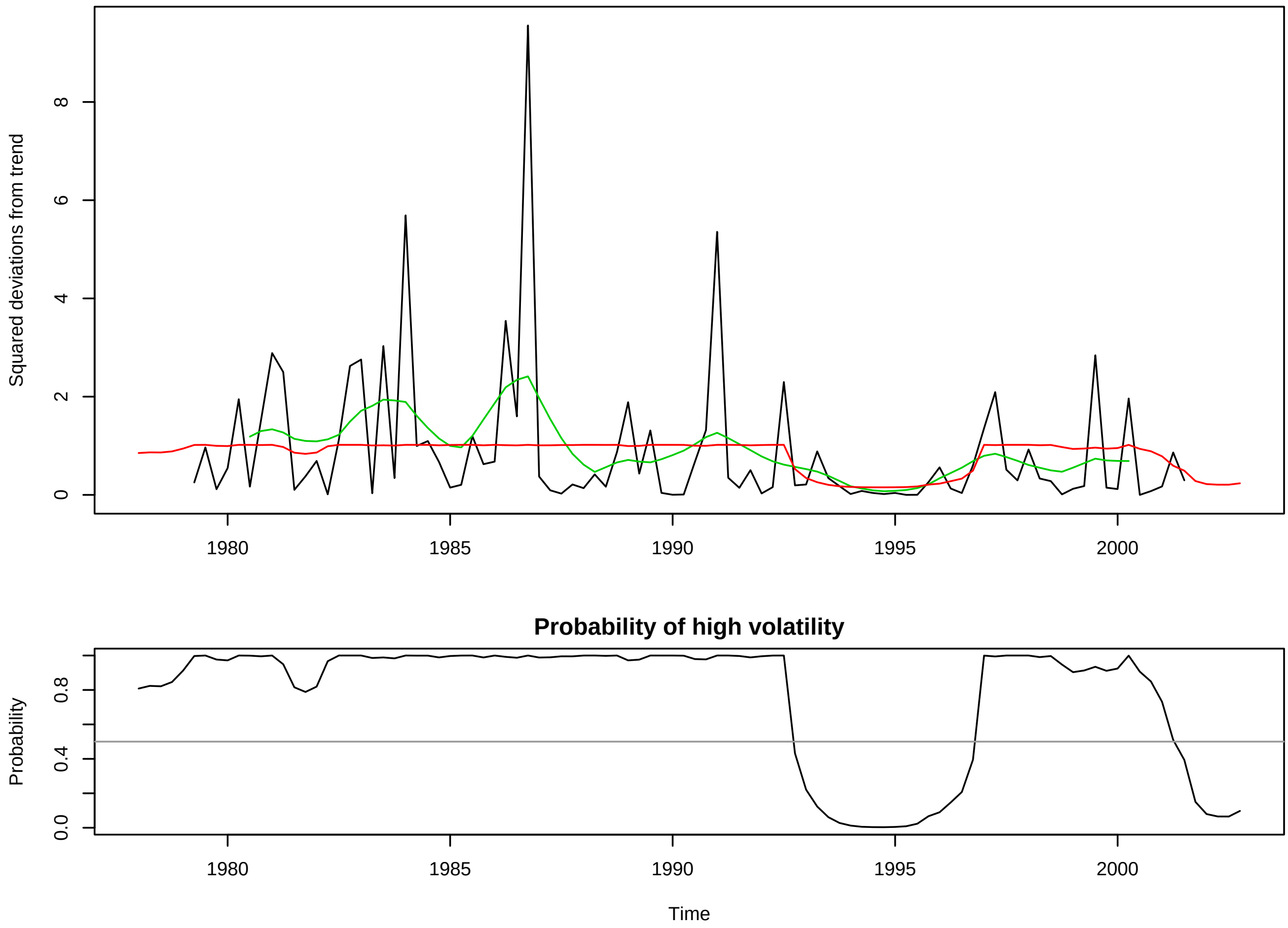
Is additional complexity of the 3–2 model justified?



Upper plot shows Y_t with HMM trend $E(\mu_{S_t}|\mathbf{Y})$ (red) and 11–quarter triangular moving average (green). The grey horizontal lines are fitted μ_j.

Lower plot shows probability of being in a high growth regime ($S_t = 1, 3, 4$) given $\mathbf{Y}$.

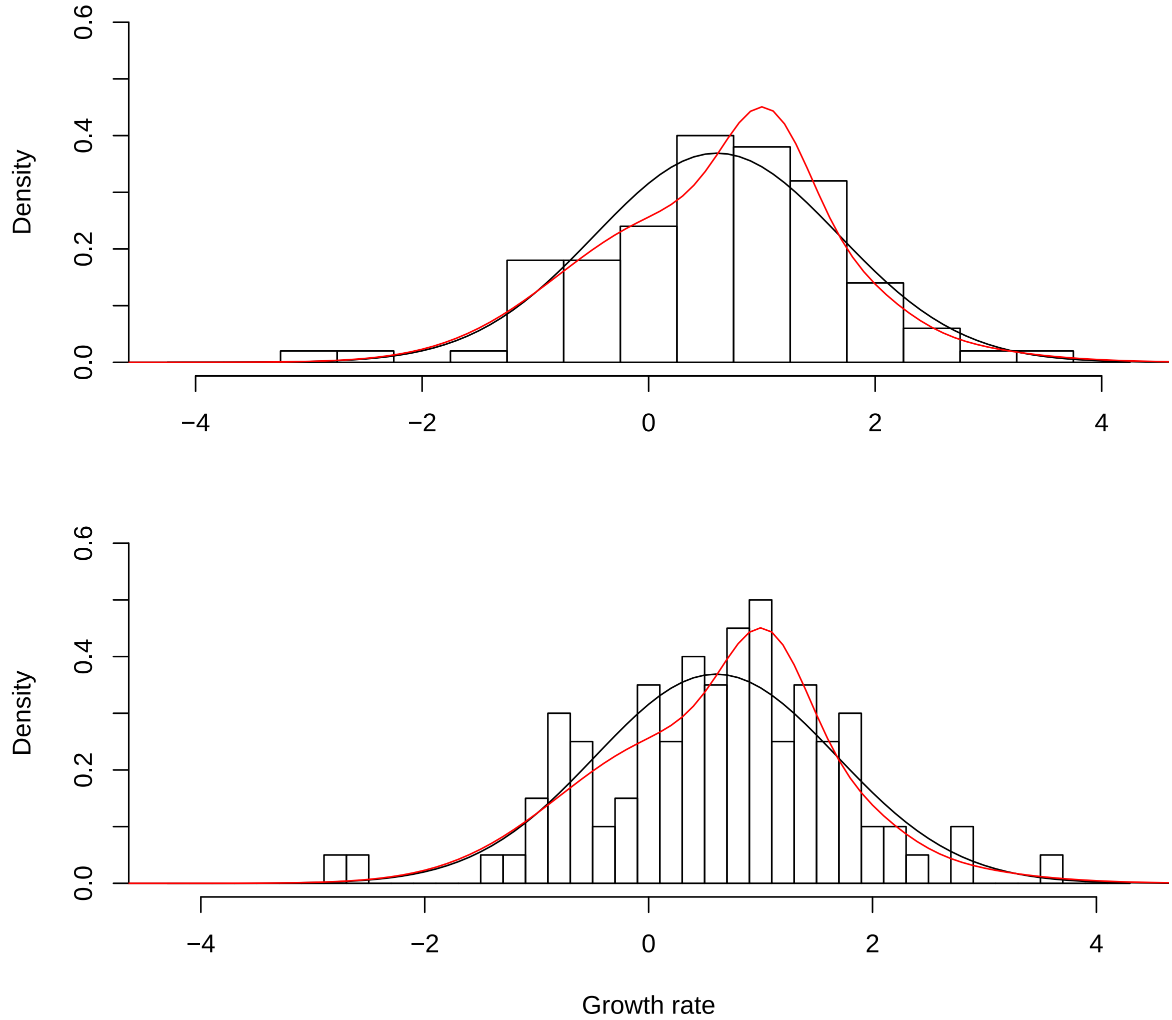
NZ GDP growth rates: 3–2 model volatility regimes



Upper plot shows squared deviations of Y_t from its moving average with HMM volatility $E(\sigma^2_{S_t}|\mathbf{Y})$ (red) and triangular 11–quarter moving variance (green).

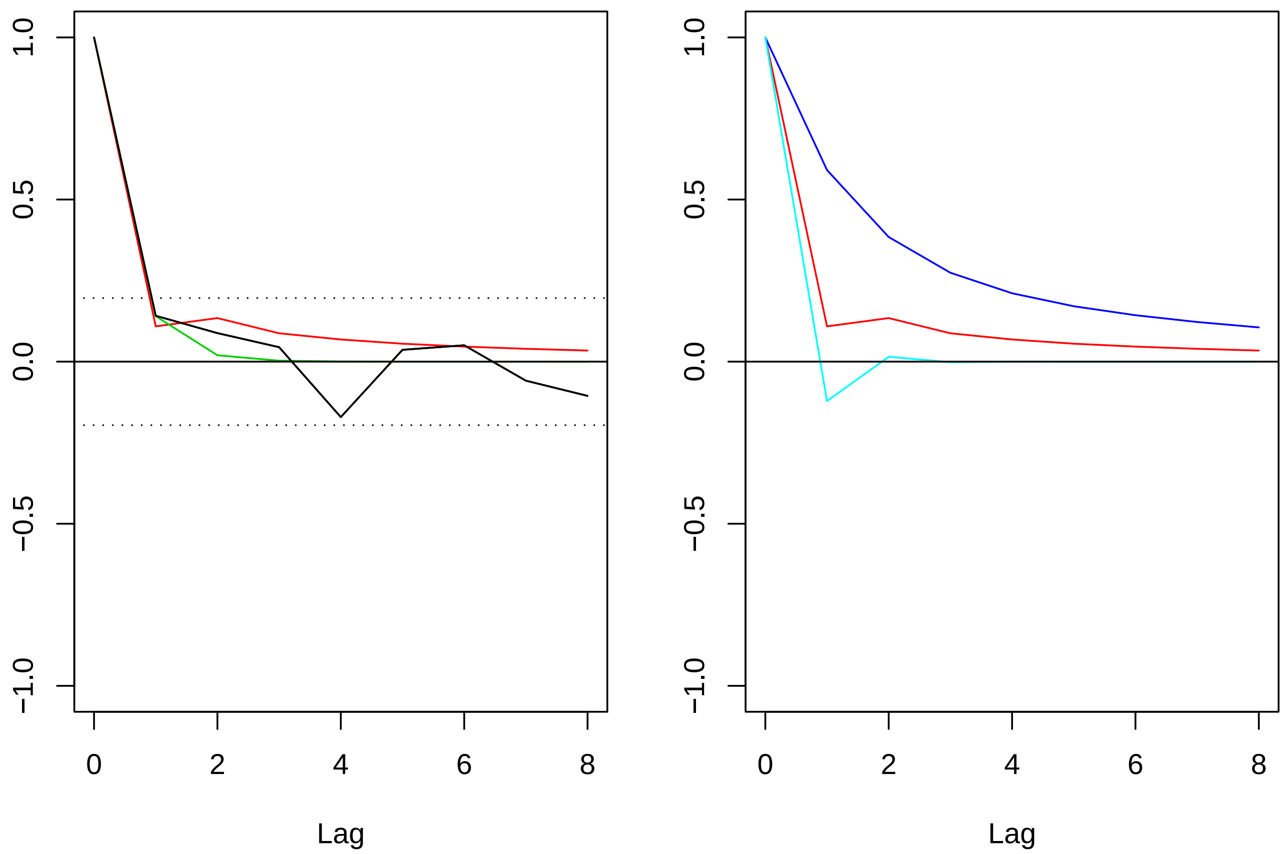
Lower plot shows probability of being in a high volatility regime ($S_t = 2, 4$) given $\mathbf{Y}$.

Histograms of NZ GDP growth rates



Histograms of Y_t at two resolutions. Best fitting marginal distributions from the Gaussian $AR(1)$ model (black) and the 3–2 model (red) are superimposed.

Autocorrelation functions for NZ GDP growth rates



Left plot shows sample acf of Y_t (black), theoretical acf for $AR(1)$ model (green) and 3–2 model (red).

For 3–2 model, right plot shows theoretical acf of Y_t (red), level μ_{S_t} (blue) and residual $\sigma_{S_t} X_t$ (cyan).

When will these modest differences lead to better (risk) forecasts for HMM model?

For future time $t{+}\nu$ and past data $\mathbf{Y}_t = (Y_t, Y_{t-1}, \ldots)$, the distribution of $Y_{t+\nu}$ is a mixture of Gaussian predictive distributions conditioned on suitable states.

Best predictor of $Y_{t+\nu}$ is a combination of optimal forecasts since

$$E(Y_{t+\nu}|\mathbf{Y}_t) = E(BP_{t+\nu}|\mathbf{Y}_t)$$

where

$$BP_{t+\nu} = \mu_{S_{t+\nu}} + \sigma_{S_{t+\nu}}\rho^{\nu}\left(\frac{Y_t - \mu_{S_t}}{\sigma_{S_t}}\right)$$

is best predictor of $Y_{t+\nu}$ given Y_t, S_t, $S_{t+\nu}$.

Linear models will capture second–order properties well, so HMM models will only provide practically useful forecast gains when factors such as distributional shape and sample paths are important.

Implies HMM models are likely to be more useful for risk forecasting than point forecasting.

One situation where HMM may do better than $AR(1)$ is forecasting turning points.

Define t to be a peak of NZ GDP L_t if

$$L_t > L_{t-2}, L_{t-1}, L_{t+1}, L_{t+2}$$

with a trough defined similarly. (BBQ algorithm of Harding and Pagan, 2002.) Since

$$Y_t = \log L_t - \log L_{t-1} \quad (t = 1, \ldots, T).$$

the peak criterion is equivalent to

$$Y_{t-1} + Y_t > 0, Y_t > 0, Y_{t+1} < 0, Y_{t+1} + Y_{t+2} < 0$$

and similarly for the trough.

Note that the last time point that can be classified is at $T-2$ (the 3rd last quarter).

Using predictive distributions of Y_t, turning points can be forecast for times $T-1$ and T using analytic formulae for the probability of a peak (trough).

How well can this be achieved?

Preliminary results only based on an incomplete simulation study.

For the two NZ GDP models:

- peaks are harder to predict that troughs
- peaks at $T-1$ and troughs at $T-1$, T are well–predicted
- peaks and troughs at other times are not well–predicted
- non–linear HMM model has better short–term prediction performance than linear AR model

More work to be done.

- **Multisite daily rainfall (Thompson et al, 2005)**

Consider a small network of K rainfall stations and

$$R_t(k) = \textit{accumulated rainfall over day } t$$

at station k. Canterbury daily data available from 1972–1997 with $K = 7$.

Aim: to provide realistic scenarios of future daily rainfall variability over diverse spatial and temporal scales (months to years). [Risk forecasting]

Need to account for seasonality, ENSO etc and build in sufficient persistence of wet and dry periods to generate suitable extremes.

Starting point: Wilks (1998) multisite rainfall simulation model used within NIWA on an operational basis for over 5 years.

Reformulated as a more general HMM rather than a simulation model.

Model

$$R_t(k) = \beta_{S_t(k)} X_t(k)$$

where

- rainfall states $S_t(k)$ form an irreducible ergodic 3 state Markov chain with

$$S_t(k) = \begin{cases} 0 & (\text{Dry at time } t) \\ 1 & (\text{Light rain at time } t) \\ 2 & (\text{Heavy rain at time } t) \end{cases}$$

- $X_t(k)$ are temporally independent exponentials with $E(X_t(k)) = 1$ and $\beta_0 = 0 < \beta_1 < \beta_2$.

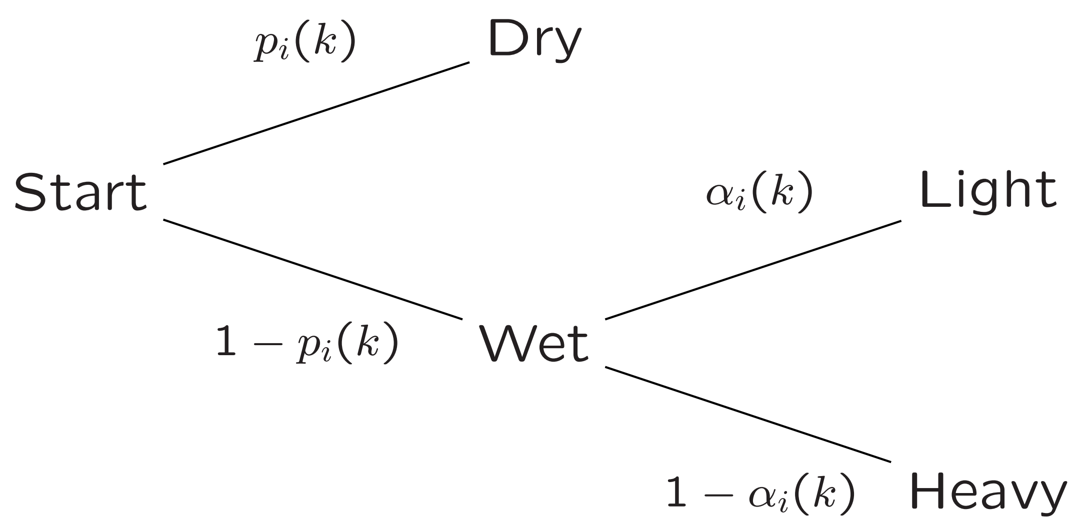
$R_t(k)$ is a mixture of two exponentials with point mass at 0.

Model contemporaneous dependence of $X_t(k)$ by

$$X_t(k) = -\log(\Phi(V_t(k)))$$

where $\Phi(.)$ is the standard Gaussian cdf, the $\mathbf{V}_t = (V_t(1), \ldots, V_t(K))$ are iid $N(\mathbf{0}, \mathbf{\Psi})$, and $\mathbf{\Psi}$ is a correlation matrix. [Meta–Gaussian copula]

For each k assume $S_t(k)$ is a stationary 3–state Markov chain with transition structure



given $S_{t-1}(k) = i$. For each location there are 6 parameters $p_i(k)$, $\alpha_i(k)$ $(i = 0, 1, 2)$.

This structural specification is more general than the Wilks (1998) model where

$$p_1(k) = p_2(k), \quad \alpha_0(k) = \alpha_1(k) = \alpha_2(k)$$

and only 3 parameters $p_0(k)$, $p_1(k)$, $\alpha_0(k)$ required for each location.

Model spatial dependence of $S_t(k)$ using

$$S_t(k) = \begin{cases} 0 & (U_t(k) \leq a_i(k)) \\ 1 & (a_i(k) < U_t(k) \leq b_i(k)) \\ 2 & (b_i(k) < U_t(k)) \end{cases}$$

when $S_{t-1}(k) = i$ with

$$a_i(k) = \mathbf{\Phi}^{-1}(p_i(k))$$
$$b_i(k) = \mathbf{\Phi}^{-1}(p_i(k) + \alpha_i(k)(1 - p_i(k))).$$

The $\mathbf{U}_t = (U_t(1), \ldots, U_t(K))$ are iid $N(\mathbf{0}, \mathbf{\Omega})$, independent of $\mathbf{V}_t$, and $\mathbf{\Omega}$ is a correlation matrix.

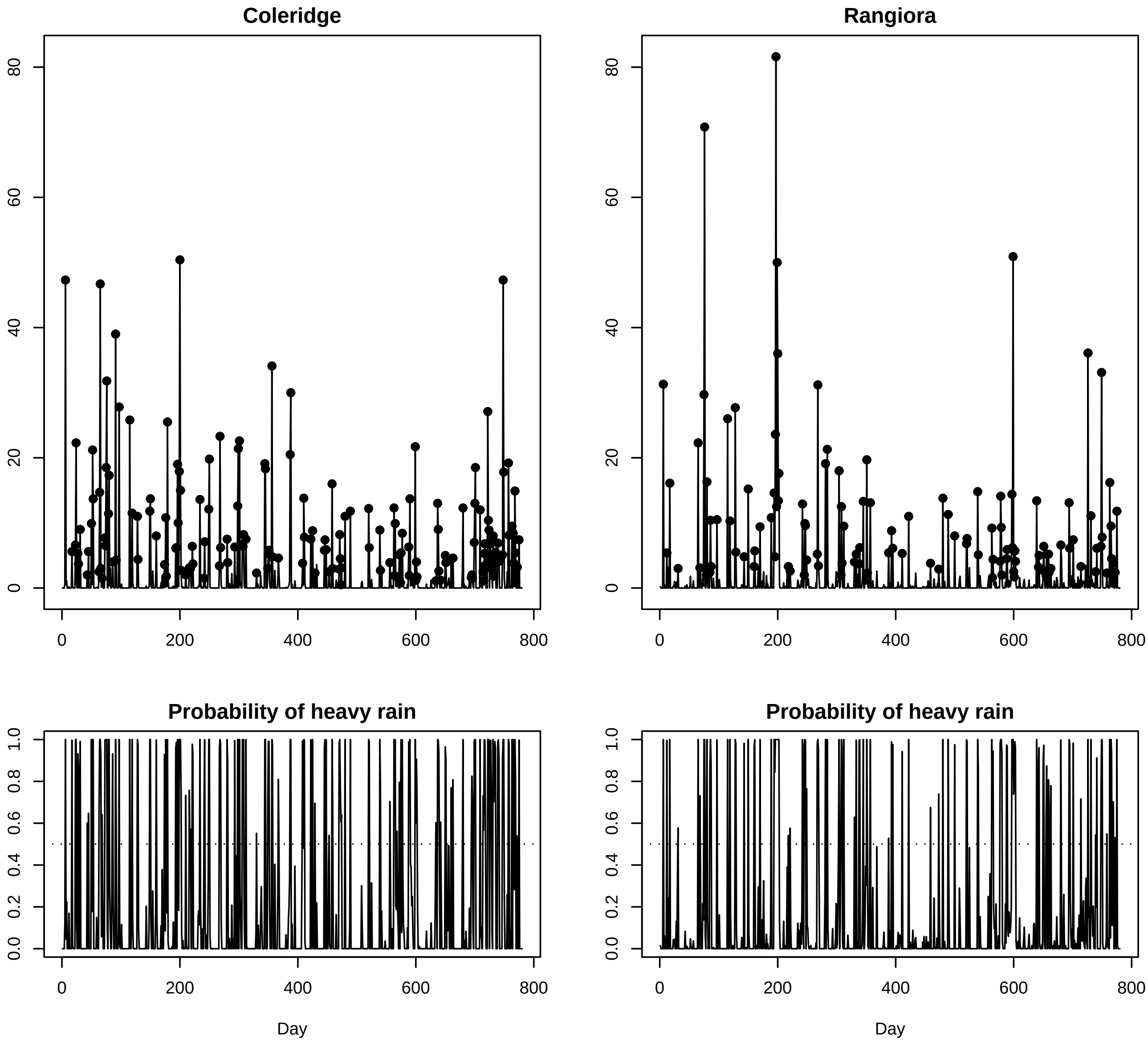
This specification, as before,

- is consistent with the (marginal) Markov chain specification;
- builds a relatively flexible joint distribution from given marginals.

Rainfall amounts $R_t(k)$ are driven by the Gaussian vector processes $\mathbf{U}_t$, $\mathbf{V}_t$ and readily simulated.

Model fitted using the usual HMM technology.

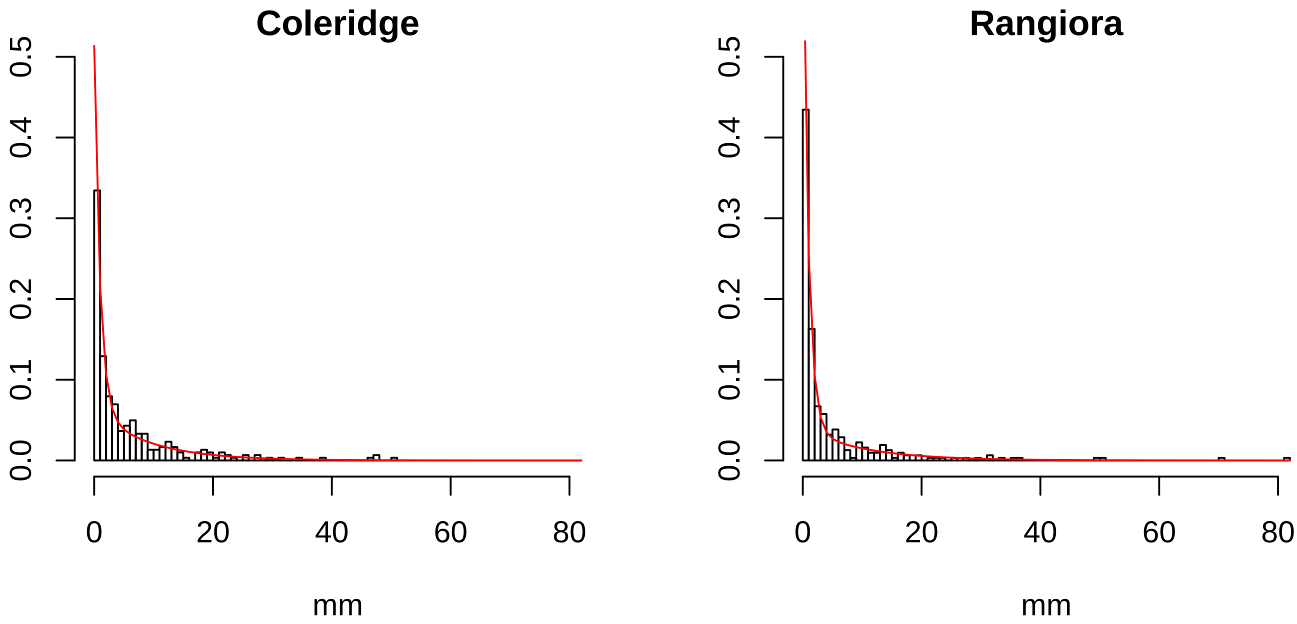
Coleridge and Rangiora daily rainfall for April



Upper plots show rainfall with superimposed points identifying rainfall classified as heavy rain.

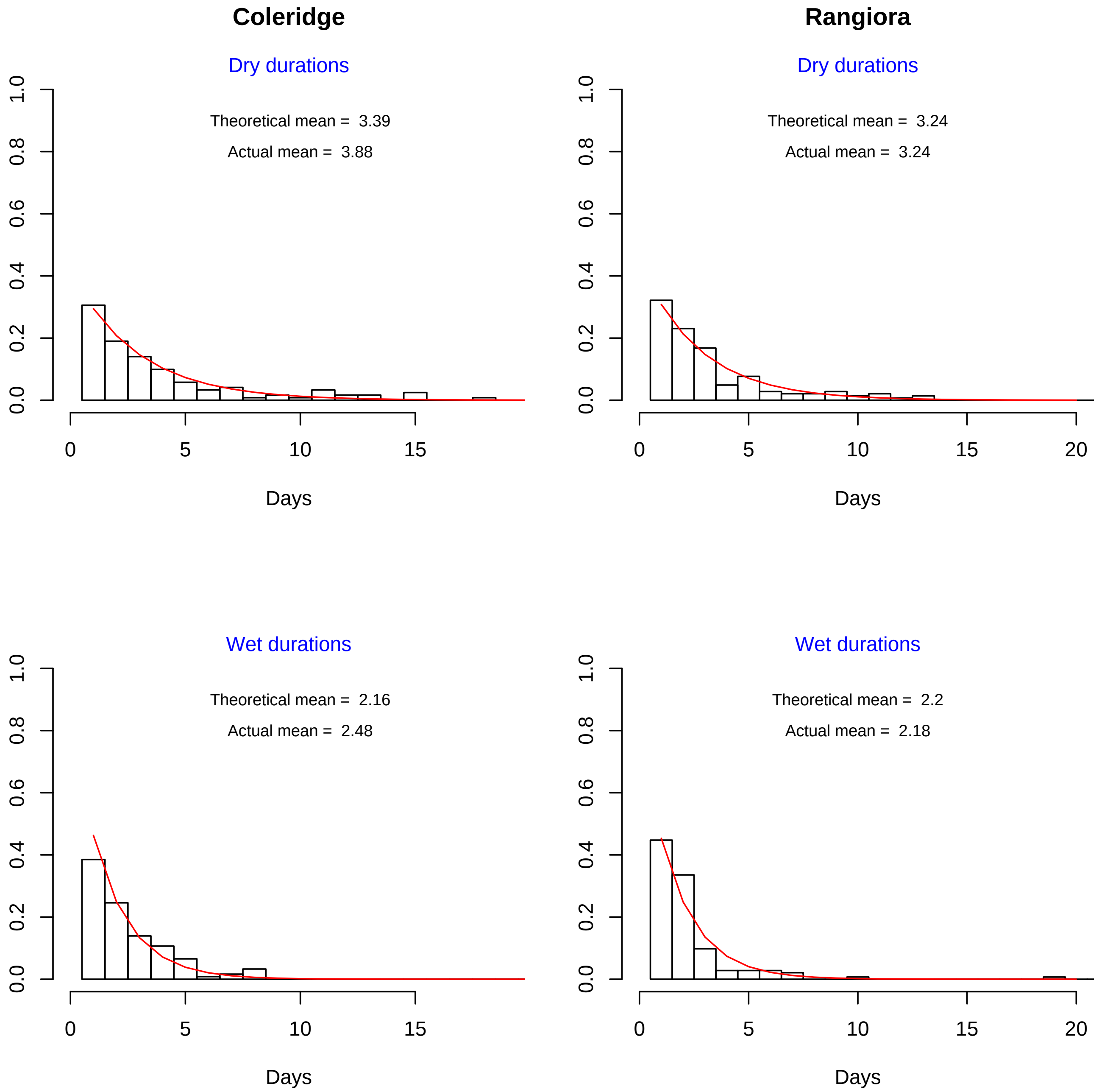
Lower plots show probability of heavy rain given the data.

Coleridge and Rangiora daily rainfall for April



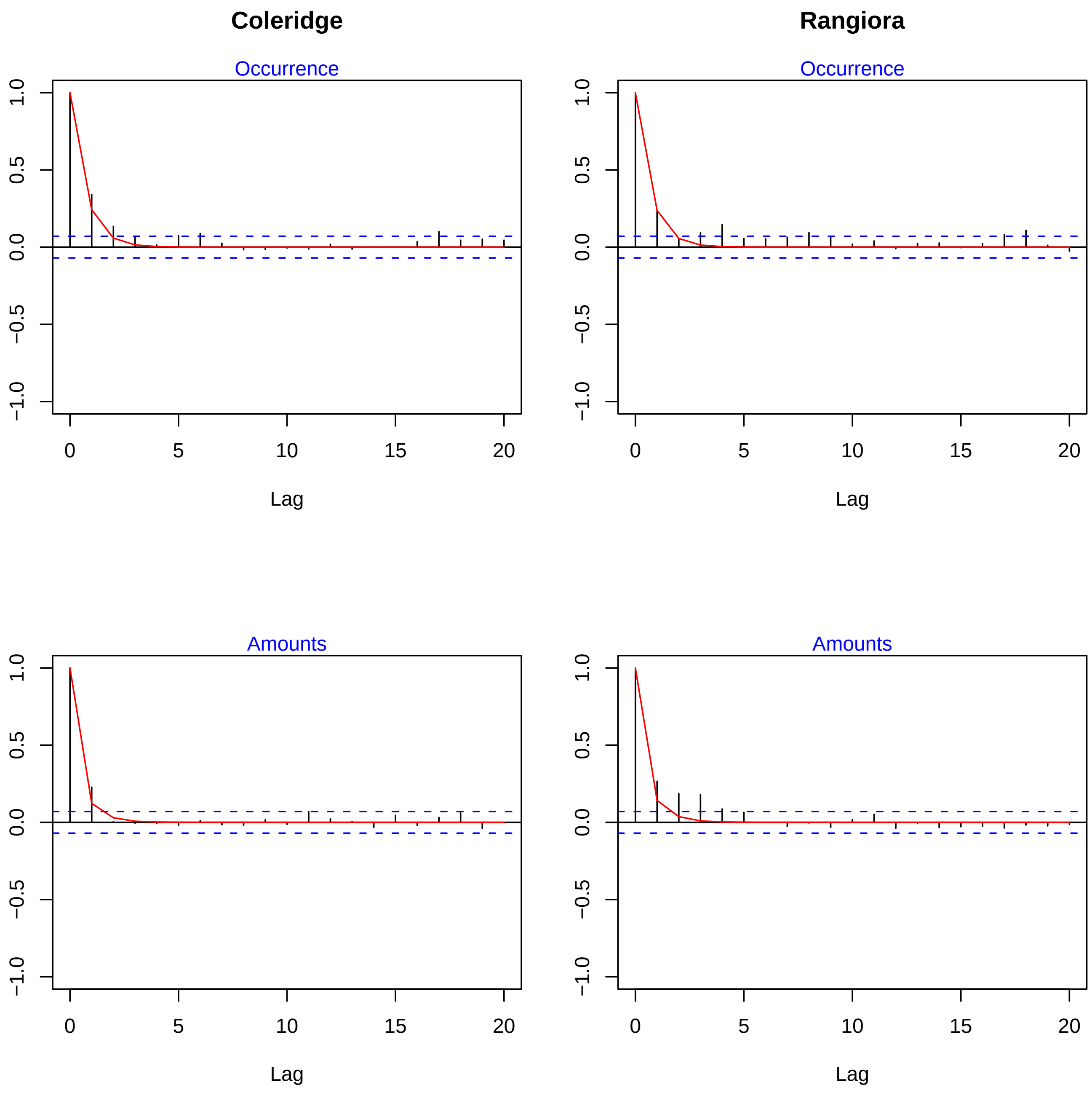
Histograms of daily rainfall on wet days with fitted exponential mixtures superimposed.

Coleridge and Rangiora daily rainfall for April



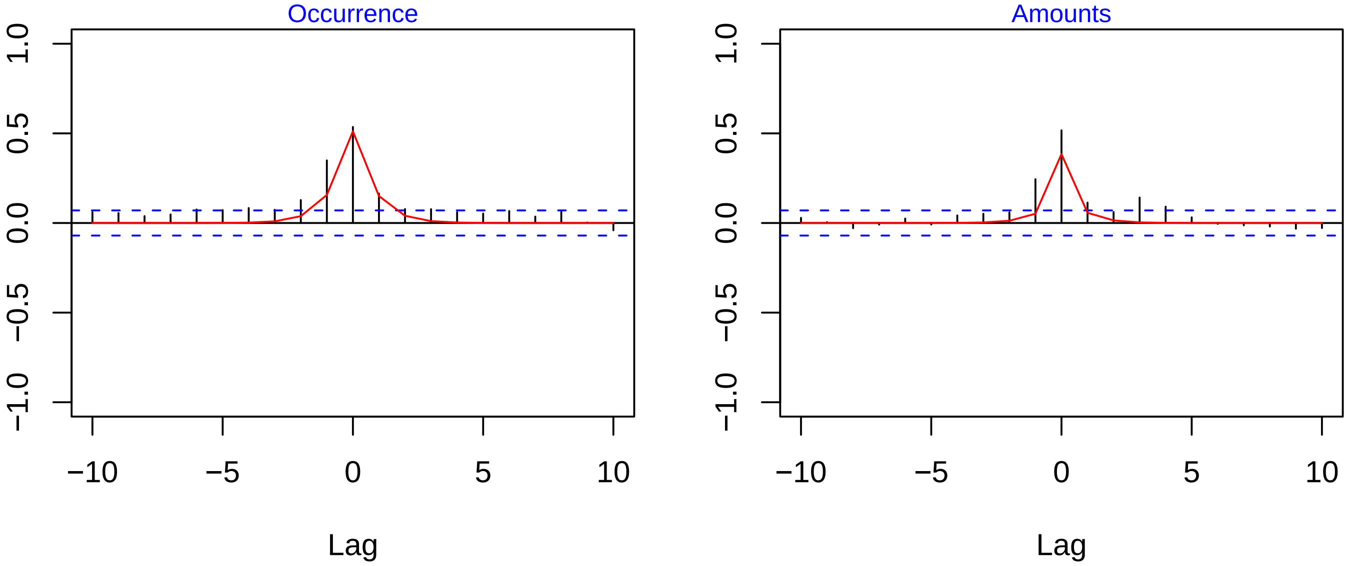
Histograms of dry and wet durations with fitted distributions superimposed.

Coleridge and Rangiora daily rainfall for April



Autocorrelation functions of rainfall occurrence and amounts with fitted autocorrelations superimposed.

Coleridge and Rangiora daily rainfall for April



Cross–correlation functions of rainfall occurrence and amounts with fitted cross–correlations superimposed.

Findings:

- AIC rarely supports Wilks model.
- Other reduced models currently being explored, with spatial homogeneity favoured in many cases.
- Rainfall distributions modelled reasonably well. [Statics]
- Dry duration distributions and cross–correlations not always well–modelled. [Dynamics]

Further developments

- Replace statistical copulas by more physically based spatial switching. [Better dynamics]
- Incorporate stochastic seasonal switching (NHMM) and account for longer–term variation.

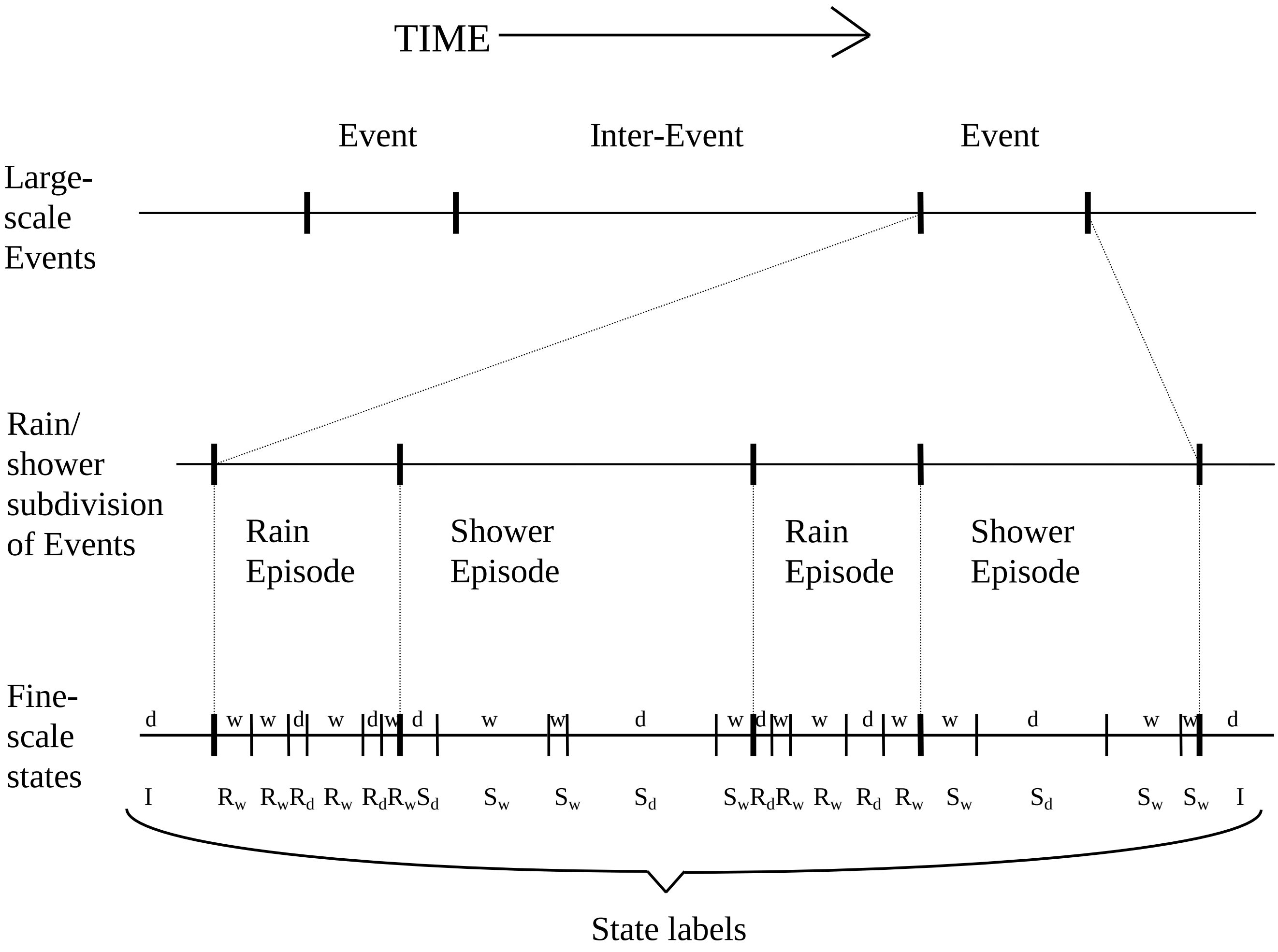
Continue the move from statistical to physical model.

- **Other**

Hidden semi–Markov model (HSMM) for breakpoint rainfall data (Sansom and Thomson, 2001)

- HSMM generalises HMM to arbitrary sojourn times (HMM has geometric sojourn times).
- Breakpoint data have very high time resolution (effectively continuous).
- Diverse time scales present in the data, from minutes to days.
- Like the daily rainfall model, dynamics are modelled by a hidden rainfall state S_t.
- HSMM fitted using EM and a variant of the HMM technology.

Breakpoint rainfall time scales



HMMs and their variants allow short–memory modelling over coarser time scales to approximate long–memory over finer time scales.

Weekly hydro inflows (Harte et al, 2004)

- Streamflow models are commonly based on periodic autoregressive moving average ($PARMA$) models.

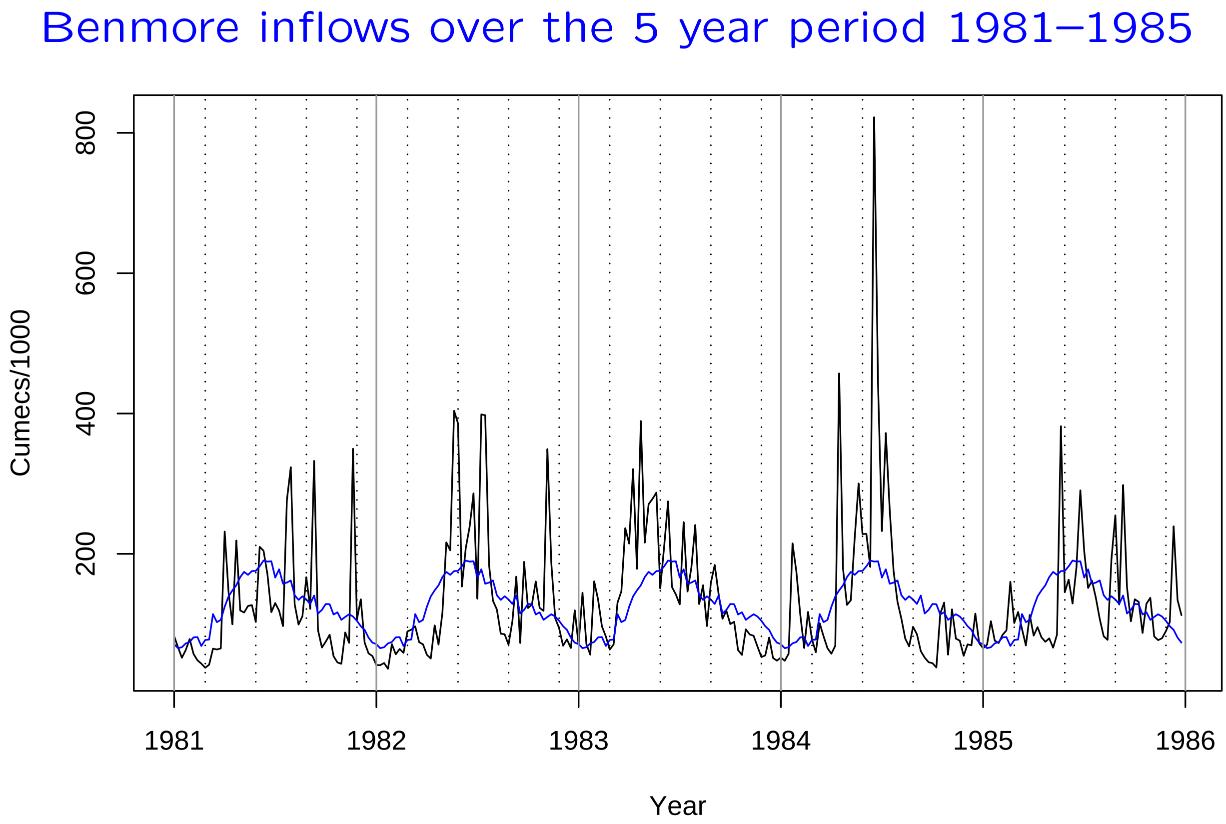
- Original $PAR(1)$ model appears to date back to Hannan (1955) who used it to model Sydney rainfall.

- Many time scales present in the NZ data.

- Weekly dynamics in NZ data dominated by evolving and episodic seasonal patterns that can switch between seasons earlier or later than expected.

Aim: to generate realistic forward realisations over seasonal to multi–year timescales that are suitable for risk forecasting, particularly of extremes.

Need to capture stochastic properties of historic inflow sequences sufficiently accurately.



Benmore inflows with mean seasonal pattern superimposed (blue).

The solid vertical grey lines mark calendar years and the dotted vertical lines mark nominal seasons.

Possible future development: to build a suitable HMM model for NZ hydro inflows that extends the basic PARMA structure.

5 Concluding remarks

HMMs have broad applicability and offer

- distributional versatility
- control of sample paths
- ability to model diverse time scales
- potential for enhanced risk forecasting
- stochastic structure to exploit

However

- care is needed with computation (few HMM packages are available)
- parsimonious (structural) hidden Markov chains may be necessary
- linear models may do just as well in situations where second moments are sufficient (e.g. point forecasting)

Major virtue? Open structure of HMM allows for more physical models which can engage statisticians, scientists, economists and clients alike in productive model development.

References

Buckle, R.A., Haugh, D. and Thomson, P.J. (2004) Markov switching models for GDP growth in a small open economy: the New Zealand experience. *Journal of Business Cycle Measurement and Analysis* **1**, 227–257.

Sansom, J. and Thomson, P.J. (2001) Fitting hidden semi–Markov models to breakpoint rainfall data. *Journal of Applied Probability* **38A**, 152–167.

Thompson, C.S., Thomson, P.J. and Zheng, X. (2005) A multisite rainfall generation model applied to New Zealand data. *NIWA Technical Report* **128**, National Institute of Water and Atmospheric Research, New Zealand.

Harte, D.S., Pickup, M. and Thomson, P.J. (2004) Stochastic models for hydro catchment inflows; an exploratory analysis. Report undertaken for the New Zealand Electricity Commission.
http://www.electricitycommission.govt.nz/opdev/modelling/hydrology/