# Time Series Analysis

## Regime Switching and State-Space Introduction

Michael Colaresi
colaresi@msu.edu

## Regime Switching

- Suppose you have a model:

$$y_t = c_1 + \phi y_{t-1} + \epsilon_t \quad (1)$$

where $\epsilon_t \sim N(0, \sigma^2)$, and this model seems adequate for $t = 1, 2, \ldots t_0$.

- However, after $t_0$, a different model holds:

$$y_t = c_2 + \phi y_{t-1} + \epsilon_t \quad (2)$$

  - This is simply a change in intercept, and thus a change in equilibrium behavior.
  - The structural change implies that the series is not stationary over the entire period.
  - However, the errors to this series are stationary:

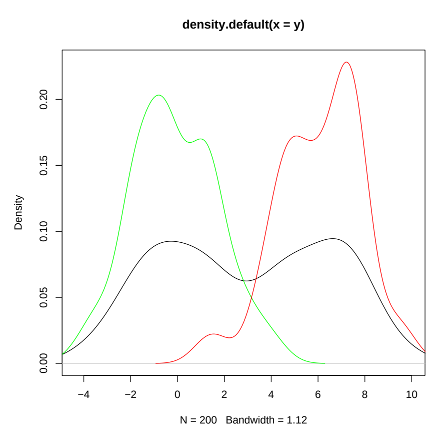
$$y_t = c_1 + c_{diff} I_t + \phi y_{t-1} + \epsilon_t$$

where $I_t = 0$ if $t \leq t_0$ and $I_t = 1$ if $t > t_0$. Therefore $c_2 = c_1 + c_{diff}$.

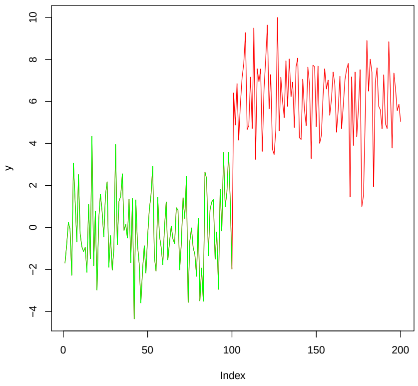
## Ignoring switch

- If we were to ignore the switch what would our data look like?
  - First case: mean shift, no dynamics.
  - Second case: mean constant, error changes.
  - Third case: Dynamics change.
    - All cases more difficult if switches occur randomly throughout data, rather than one or a small number of switches.

## Mean Shift I



## Mean Shift II




## Variance Switching

- Suppose you have a model:

$$y_t = c_1 + \phi y_{t-1} + \epsilon_{1t} \tag{3}$$

where $\epsilon_{1t} \sim N(0, \sigma_1^2)$, and this model seems adequate for $t = 1, 2, \ldots t_0$.

- However, after $t_0$, a different model holds:

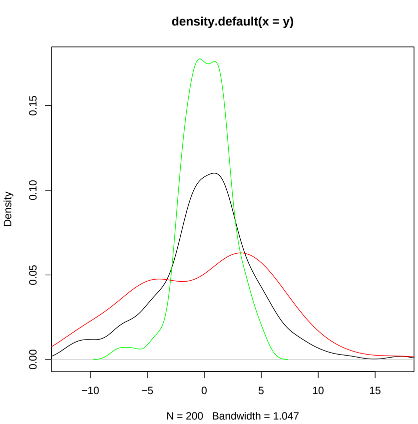
$$y_t = c_1 + \phi y_{t-1} + \epsilon_{2t} \tag{4}$$

where where $\epsilon_{1t} \sim N(0, \sigma_2^2)$; and $\sigma_1 \neq \sigma_2$

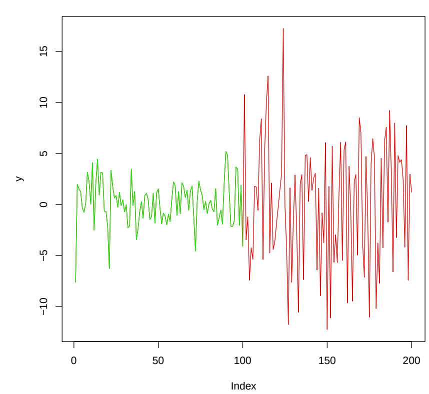
  - This is simply a change in error variance.
  - The structural change implies that the series is not stationary over the entire period.
  - The errors, conditional on splitting the sample at $t_0$, are stationary:

$$\begin{aligned} y_t &= c_1 + \phi y_{t-1} + \epsilon_{1t}, \forall t \leq t_0 \\ y_t &= c_1 + \phi y_{t-1} + \epsilon_{2t}, \forall t > t_0 \end{aligned}$$

## Variance Shift I




## Variance Shift II



## Dynamics Switching

- Suppose you have a model:

$$y_t = c_1 + \phi_1 y_{t-1} + \epsilon_t \tag{5}$$

where $\epsilon_t \sim N(0, \sigma^2)$, and this model seems adequate for $t = 1, 2, \ldots t_0$.

- However, after $t_0$, a different model holds:

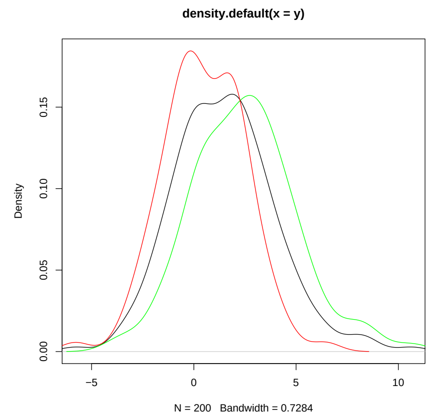
$$y_t = c_1 + \phi_2 y_{t-1} + \epsilon_t \tag{6}$$

  - This is a change in dynamics, and thus a change in equilibrium behavior.
  - The structural change implies that the series is not stationary over the entire period.
  - However, the errors to this series are stationary, such that:

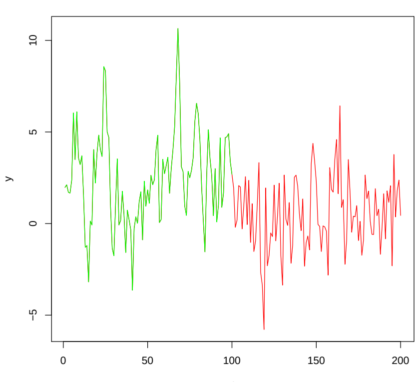
$$y_t = c_1 + \phi_1 y_{t-1} + \phi_I y_{t-1} I_t + \epsilon_t$$

where $I_t = 0$ if $t \leq t_0$ and $I_t = 1$ if $t > t_0$. Therefore $\phi_2 = \phi_1 + \phi_I$.

## Dynamics Shift I




## Dynamics Shift II




## TAR models

Regimes do not have to be subset by time.

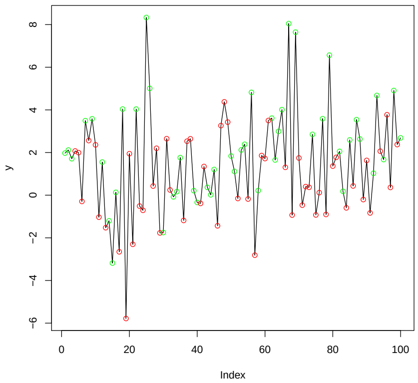
Can be purely stochastic–picking/sampling from a hat.

The threshold autoregression model attempts to pick up different "states" that depend on the previous value of the dependent variables (only). Two state example:

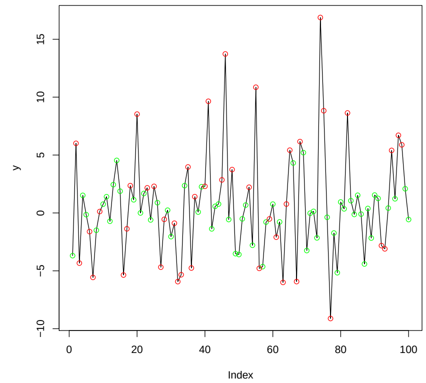
$$\begin{aligned} y_t &= \alpha^{(1)} + \Sigma_{j=1}^{p} \phi_{t-j}^{(1)} y_{t-j} + w_t^{(1)}, y_{t-1} < R \\ y_t &= \alpha^{(2)} + \Sigma_{j=1}^{p} \phi_{t-j}^{(2)} y_{t-j} + w_t^{(2)}, y_{t-1} \geq R \end{aligned}$$

- Assumes $r$(number of cut points) and $R_r$ (location of cutpoints) are known.
- Assumes $p$ is known.
- Assumes $w_t^{(r)}$ are independent.
- Estimate R separate models.
- Including $X$s are no problem.
- Developed by Tong(1983)

## Dynamics in a Blender




## Error Variances in a Blender




## Bait and Switch

- There are a few important question to ask about your data if you think more than one regime is present.
  - Obvious: What changes? How many different states/regimes?
  - Less obvious: How "enduring" are changes in states.$\star$

    *Absorbing* $\Rightarrow$ *Random*

## Discrete Markov Process

Suppose that $X_n; n = 0, 1, 2, \dots$ is a stochastic process.

- Suppose also that $X_n$ can only take on values in the set $1, 2, \dots, m$.
- Think of these values as $m$ "states" of a system.
  - Economic good times vs. bad times.
  - times of high international tension vs. peace.

Markov process: Aka Markov chain is a stochastic process where the probability of the next (future) state given the current state and the entire past depends only on the current state.

$$P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

## Transition Matrix

Markov Process/Chain is a constraint on the way the past state can influence the present state (only through the present).

We can then define the transition probabilities between the $m$ states with a $m \times m$ transition matrix, $\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & \cdots & \cdots & p_{mm} \end{bmatrix}$

Which we can read as $p_{ij}$ is the probability of moving from state $j$ to state $i$, which is $P(X_n = x_i | X_{n-1} = x_j)$.

## Coin flip example

- Coin has two states, $m = 2$, heads (1) or tails (2).
- Markov chain representation for each would be:

$$\begin{aligned} P(X_n = 1 | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots) &= P(X_n = 1 | X_{n-1} = x_{n-1}) \\ &= p \end{aligned}$$

The probability of seeing tails is just 1 minus the probability of heads,

$$1 - P(X_n = 1 | X_{n-1} = x_{n-1}) = 1 - p \tag{7}$$

Transition matrix:

$$\left[\begin{array}{c|cc} & Heads(t-1) & Tails(t-1) \\ \hline Heads(t) & p & p \\ Tails(t) & 1-p & 1-p \end{array}\right]$$

## Absorbing vs. Stochastic

The values within the transition matrix tell us how "sticky" the states are.

- What does a value of 1 on the diagonal imply?
- What does it imply if the diagonal value for one row equals all other values in the row (as in the coin flip example)?

## Markov Switching Models

- Markov process representation are useful because we can look at regime changes as stochastic rather than deterministic.
- Think about the previous deterministic examples (figures).
  - Where the shifts in mean, error or dynamics perfectly predictable?
  - Models that "hard-code" cutpoints in the model (on y or in time) assume that the change was deterministic and could have been forecasted perfectly.
- What we want is to model both the dynamics of the states as well as the transition between states.

## Markov Swiching Regression Model

$$\begin{aligned} y_t &= c_1 + \phi_1 y_{t-1} + \epsilon_{1t}, if\ s_t = 0 \\ y_t &= c_2 + \phi_2 y_{t-1} + \epsilon_{2t}, if\ s_t = 1 \\ Pr(s_t = i | s_{t-1} = j, \ldots) &= Pr(s_t = i | s_{t-1} = j) \end{aligned}$$

- We can then parameterize the last equality with a logit, probit or some other specification.
- Can then maximize the likelihood, conditional on initial parameters for the three equations.

## Markov Review

- Remember, that a Markov process is some stochastic process that has a current state that only depends on yesterday's state, not anything that came before.

  $P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$

- If there are $m$ states, that means we can define an $m \times m$ transition matrix that will hold all of the conditional probabilities

  $$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & \cdots & \cdots & p_{mm} \end{bmatrix}$$

  where, $p_{ij} = P(X_n = x_i | X_{n-1} = x_j)$
  - $\frac{\Sigma_{j=1}^{m} p_{ij}}{m}, \ \forall\ i \leq m \Leftarrow p(X_t = 1)$
  - $\Sigma_{i=1}^{m} = p_{ij} = 1, \ \forall\ j \leq m$

## Markov Memory

- The idea that a process only depends on one time period previous seems restrictive (especially for a time series class).
- However, by redefining the states, we can recapture the simplicity of the markov process, even when there is some higher order dependence.
  - My state today depends on not only what state I was in yesterday, but also the day before.

## State of Transition

Lets say that some process $X$ now depends on yesterday and today, and lets say $m = 2$ for simplicity.

- Think about this as two states that are a "healthy economy" and a "depressed economy".
- A simple first order markov model would say that we could define a model for the probability of being in a healthy or depressed economy tomorrow by only knowing the what state we are in currently.
- But what if there is "backsliding", you may come out of a depression for one time period, but have a high probability of sliding back into depresssion the next year.
- On the other hand, if you can stay healthy for two time periods consecutively, the probability of reverting back to depression decreases.

## Higher-order Markov Processes

What we are doing can be seen in matrix form as:

$$\Pi_t = T\Pi t - 1 \tag{8}$$

- $\Pi_t$ is a column state vector at time $t$. Tells you what proportion of the population is in each state at time $t$ $(m \times 1)$.
- $T$ is simply our transition matrix $(m \times m) \leftarrow$ rows tell you where you are going (i) and the columns (j) tell you the previous state (coming from).

## Continuing

This is, $\begin{bmatrix} \Pi_{1,t} \\ \Pi_{2,t} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} \Pi_{1,t-1} \\ \Pi_{2,t-1} \end{bmatrix}$

- This all means that,

$$\begin{bmatrix} \Pi_{1,t} \\ \Pi_{2,t} \end{bmatrix} = \begin{bmatrix} (p_{1\leftarrow 1}\Pi_1, t-1)+ & (p_{1\leftarrow 2}\Pi_{2,t-1}) \\ (p_{2\leftarrow 1}\Pi_1, t-1)+ & (p_{2\leftarrow 2}\Pi_{2,t-1}) \end{bmatrix}$$

## Depression, ignoring backsliding

- start with $Pi_{t-1} = [0, 1]'$, all depressed.
- $T = \begin{bmatrix} 1 & .2 \\ 0 & .8 \end{bmatrix}$
- What are the transitions, at time $t = 1, 2, 3$

## Dimensions of the System

If we want to capture second order dependence (yesterday and the day before), we just redefine states $s$ which had $m$ distinct states, as states $s*$ which will have $m^2$ states.

- Previous states ($s$): healthy, depressed.
- New states ($s*$):
  1. healthy(t-1), healthy(t-2)
  2. healthy(t-1), depressed(t-2)
  3. depressed(t-1), healthy(t-2)
  4. depressed(t-1), depressed(t-2)
- Thus we have a new $T*$ transition matrix that is $m^2 \times m^2$.

## Matrix form

$$\begin{bmatrix} \Pi_{1,t} \\ \Pi_{2,t} \\ \Pi_{3,t} \\ \Pi_{4,t} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{23} & p_{24} \\ p_{31} & p_{32} & 0 & 0 \\ 0 & 0 & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} \Pi_{1,t-1} \\ \Pi_{2,t-1} \\ \Pi_{3,t-1} \\ \Pi_{4,t-1} \end{bmatrix}$$

- Note that if $p_{i,j} = p_{i,j+1}; \forall\ i,j|i$ we are back to first order markov process.
- Intuition - 8 parameters in this model $m^3$, but 4 parameters in first order simple markov process model.

## Second order example

- start with $Pi_{t-1} = [0,0,0,1]'$, all depressed, and have been depressed for at least one time period previous.
- $T = \begin{bmatrix} 1 & .8 & 0 & 0 \\ 0 & 0 & .6 & .2 \\ 0 & .2 & 0 & 0 \\ 0 & 0 & .4 & .8 \end{bmatrix}$
- What do these mean?
- How would you go about computing proportions?

## Fake Dynamic Logit

- You will see some examples of people using logit, probit, cloglog models as "Markov Transition models".
- For example, you might have one logit equation measuring the probability of an autocracy becoming democratic (or staying autocratic) and another logit measuring whether whether democratic countries become autocratic (or stay democratic).
  - Logits are estimated as if they are independent.
  - Think about what this is.
  - You are just creating m-equations, and parametrizing (with x's) each row of $\Pi_t$.
  - Allows covariates to have a different effect depending on the previous value of the state (where you are democratic country last year or a republican country last year).
  - Now you know that these are merely restricted higher order Markov models, where the more distant past (previous to one year) does not influence the probability of changing states.

## Muppets in (State-) Space

The general state-space model has two equations,

$$\begin{aligned} y_t &= F\theta_t + \upsilon_t \\ \theta_t &= G\theta_{t-1} + \omega_t \end{aligned}$$

where, $\upsilon_t \sim N(0, V_t)$ and
$\omega_t \sim N(0, W_t)$.

| Parameter | Definition | Notes |
|---|---|---|
| $y_t$ | Observation | Can be multivariate |
| $\theta_t$ | Unobserved state | Can be multivariate |
| $F$ | observation matrix | includes known and unknown parameters |
| $G$ | transition matrix | includes known an unknown parameters |
| $\upsilon_t$ | observation noise | normal assumptions, $W_t$ is covariance |
| $\omega_t$ | structural noise | normal assumptions, $V_t$ is covariance |

## Matrix dimensions: Mind Your Ps and Qs

$p$ is the number of " unobserved states", while $q$ is the number of "measurements/observations" at time $t$.
These define the dimension of the system.

1. $p = q$
2. $p < q$, dynamic factor model.
3. $p > q$, decomposition of variance

- $\theta t$ is a $p \times 1$ vector.
- $G$ is $p \times p$.
- $\upsilon_t$ is $p \times 1$.
- $y_t$ is a $q \times 1$ vector.
- $F$ is $q \times p$.
- $omega_t$ is $q \times 1$

## Matrix Dimensions II

$$\theta_t = \begin{bmatrix} \theta_{t1} \\ \theta_{t2} \\ \vdots \\ \theta_{tp} \end{bmatrix}, G = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & \cdots & \cdots & a_{pp} \end{bmatrix}, y_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tq} \end{bmatrix},$$

$$F = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & \cdots & \cdots & a_{qp} \end{bmatrix}$$

## An Example: Presidential Approval

3 polls ($q = 3$), 1 "state" of approval ($p = 1$). $\theta_t = \begin{bmatrix} \theta_{t1} \end{bmatrix}$, $G = \begin{bmatrix} a_{11} \end{bmatrix}$, $y_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix}$, $F = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$

## Multivariate Series

The model above only includes one lag for the state vector. This can be generalized replacing the $p \times 1$ state vector with a $pm \times 1$ state vector (m is the number of lags for state equation).

Now, $\theta_t = \begin{bmatrix} \theta_t \\ \theta_{t-1} \\ \vdots \\ \theta_{t-m+1} \end{bmatrix}$ and each element is itself a matrix. This is known as an array.

New transition matrix ($G$) is now ($pm \times pm$),

$$\begin{bmatrix} G_1 & G_2 & \cdots & G_{m-1} & G_m \\ I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}$$

each $G_i$ is a $p \times p$ matrix, and $I$ and 0 are $p \times p$ identity or zero matrices. So you have $m$ number of $ptimesp$ matrices.

The New observation matrix is now $q \times pm$, $= [F|0|\cdots|0]$, where $F$ is $q \times p$ and the zeros are $q \times p$ zero matrices.

## Example with 2 lags

Think about previous presidential approval example, now $m = 2$ and so dimensions of state equation will change from $p \times 1$ (1,1) to $pm \times 1$ (2,1). Transition matrix will now be (2,2) not (1,1)

$\theta = \begin{bmatrix} \theta_t \\ \theta_{t-1} \end{bmatrix}$, $G = \begin{bmatrix} a_{11} & a_{12} \\ 0 & 1 \end{bmatrix}$, $y_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix}$, $F = \begin{bmatrix} b_{11} & 0 \\ b_{21} & 0 \\ b_{31} & 0 \end{bmatrix}$

## Example Continued

which we can write: $\begin{bmatrix} \theta_t \\ \theta_{t-1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \theta_{t-2} \end{bmatrix} + \upsilon_t$

and

$\begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 \\ b_{21} & 0 \\ b_{31} & 0 \end{bmatrix} \begin{bmatrix} \theta_t \\ \theta_{t-1} \end{bmatrix}$

## Generality

- Almost any constant parameter model can be written in state space form.
- Often there are more than 1 state space representations of the same underlying model (more than one way to skin a duck).
- Major benefit: Parameter dynamics.
  - It is natural in the state-space formulation to have parameters drift over time
  - In this way you could have a process where the AR parameters change as a function of time (ARIMA breaks down here).

## Kalman Filter

When you design a state space model, whatever the measurements, you are interested in the unobserved states.

- The Kalman filter is an algorithm to optimally estimate the underlying states of system of equations, given all the information up to and including time t.
- This is analogous to least squares for regression.
- Keep in mind a few things:
  - Filter: Use information up to and including time t.
  - Smoother: Include ALL information including observations AFTER time time (condition state estimation on the whole sample).
  - Prediction: Use information from past to predict present and future.
- Filter is the building block of smoother and prediction.

## Workflow for State Space Models

- Design state space formulation.
- identify unknown parameters.
- Estimate unknown parameters using ML, EM algorithm, or MCMC techniques.
- Use Kalman filter, with the computed parameter estimates to capture the unobserved states.
- Use the Kalman filter results to get the Kalman smoothed results (update state-estimates with whole sample).
- Publish, get tenure, join AARP.

## Lingo

Frequentists call these models "State-space models"

- Missile tracking.
- Thank you NASA

Bayesians call these models "Dynamic Linear Models (DLMs)"

- The dynamics are in the parameters
  - $\theta_t = G\theta_{t-1}) + \omega_t$, instead of,
  - $\theta_t = \theta_{t-1}$, where $G = I$ and $\omega_t = 0$
  - Does bayesian inference make more sense when parameters are moving.

## Why You Should Care

- Jackman's work on Presidential Approval and Votes (2006)
- Jackman's work on Dynamic Logit (2005)
- Mark Pickup's work on Economic Voting in Britain (2006)
- PEWMA and PAR (2002)
- Markov Switching BVAR- Brandt and Freeman, the next generation. (SOON?)
- Quinn, et al Topic Coding Model. (2006)
- Change-point models - ? , now in MCMCpack. (2006)
- Time-varying parameters
  - Stimson's measure of mood
  - DYMIMIC model of the Macropolity (McKeown, Erickson and Stimson)
  - Kalman Filter on systemic democracy by Sara Mitchell

## R Programs

- MCMCpack - Changepoint
- StructTS - univariate SS
- DSE - multivariate SS, hard to use
- DLM - Bayesian, multivariate, medium difficulty level.
- SSPIR - multivariate SS, includes some nonlinear functions.

Stata 12/13 does some of this with sspace.