LECTURE ON
THE MARKOV SWITCHING MODEL

CHUNG-MING KUAN

Department of Finance & CRETA
National Taiwan University

May 18, 2010
Lecture Outline

1. Introduction

2. MS Model of Conditional Mean

3. Model Estimation
   - Quasi-Maximum Likelihood Estimation
   - Estimation via Gibbs Sampling

4. Hypothesis Testing
   - Testing for Switching Parameters
   - Testing Other Hypotheses

5. Application: Taiwan’s Business Cycles
Lecture Outline (cont’d)

6 MS Model of Conditional Variance
   - Switching ARCH Models
   - Switching GARCH Models

7 MS Model of Conditional Mean and Variance

8 Application: Taiwan’s Short Term Interest Rates

9 Concluding Remarks
Time Series Models

- Linear models for conditional mean: AR, MA, ARMA, and ARMAX
- Models for conditional variance: ARCH, GARCH and their variants
- Limitations of some nonlinear models
  - Not easy to implement: Numerical search, local minimum
  - Specific for certain nonlinear patterns, such as level shift, asymmetry, volatility clustering
Markov Switching (MS) Model

- **MS model of conditional mean** (Hamilton, 1989 and 1994) and conditional variance (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996)
  - Multiple structures (equations) for conditional mean and conditional variance
  - Switching mechanism governed by a Markovian state variable
- **Features**
  - Characterizing distinct (mean or variance) patterns over time
  - More flexible than models with structural changes
  - Allowing for regime persistence (cf. random switching model)
A generic model with two structures at different levels:

\[
  z_t = \begin{cases} 
  \alpha_0 + \beta z_{t-1} + \varepsilon_t, & s_t = 0, \\
  \alpha_0 + \alpha_1 + \beta z_{t-1} + \varepsilon_t, & s_t = 1,
\end{cases}
\]

where \(|\beta| < 1\) and \(s_t = 1, 0\) is a state variable. Some examples:

- **Model with a single structural change**: \(s_t = 0\) for \(t = 1, \ldots, \tau_0\) and \(s_t = 1\) for \(t = \tau_0 + 1, \ldots, T\)

- **Random switching model**: \(s_t\) are independent Bernoulli random variables, Quandt (1972)

- **Threshold AR model**: \(s_t\) is the indicator variable \(1\{\lambda_t \leq c\}\)
Let $s_t$ be an unobservable state variable governed by a first order Markov chain with the transition matrix:

\[
P = \begin{bmatrix}
\Pr(s_t = 0 \mid s_{t-1} = 0) & \Pr(s_t = 1 \mid s_{t-1} = 0) \\
\Pr(s_t = 0 \mid s_{t-1} = 1) & \Pr(s_t = 1 \mid s_{t-1} = 1)
\end{bmatrix}
= \begin{bmatrix}
p_{00} & p_{01} \\
p_{10} & p_{11}
\end{bmatrix},
\]

- $z_t$ are jointly determined by $\varepsilon_t$ and $s_t$.
- The Markovian $s_t$ variables result in random and frequent changes.
- The persistence of each regime depends on the transition probabilities.
- Regime classification is probabilistic and determined by data.
Extensions

- AR($k$) model with a switching intercept:
  \[ z_t = \alpha_0 + \alpha_1 s_t + \beta_1 z_{t-1} + \cdots + \beta_k z_{t-k} + \varepsilon_t. \]

- VAR (vector autoregressive) model with switching intercepts:
  \[ z_t = \alpha_0 + \alpha_1 s_t + B_1 z_{t-1} + \cdots + B_k z_{t-k} + \varepsilon_t. \]

- Multiple states: $s_t$ assumes $m > 2$ values.
- Dependence on current and past state variables:
  \[ \tilde{z}_t = \beta_1 \tilde{z}_{t-1} + \cdots + \beta_k \tilde{z}_{t-k} + \varepsilon_t, \]
  where $\tilde{z}_t = z_t - \alpha_0 - \alpha_1 s_t$.
- Transition probability as a function of exogenous variables
When a unit root is present in $y_t$ such that $\Delta y_t = z_t$, we can write

$$y_t = \left( \alpha_0 t + \alpha_1 \sum_{i=1}^{t} s_i \right) + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} + \sum_{i=1}^{t} \varepsilon_t.$$  

**Figure:** The Markov trend function with $\alpha_1 > 0$ (left) and $\alpha_1 < 0$ (right).
Quasi-Maximum Likelihood Estimation

- The model parameters: \( \theta = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_k, \sigma^2_{\varepsilon}, p_{00}, p_{11})' \).
- Optimal forecasts of \( s_t = i \) (\( i = 0, 1 \)) based on different information sets:
  - **Prediction** probabilities: \( \mathbb{P}(s_t = i \mid Z^{t-1}; \theta) \), with \( Z^{t-1} = \{z_{t-1}, \ldots, z_1\} \)
  - **Filtering** probabilities: \( \mathbb{P}(s_t = i \mid Z^t; \theta) \)
  - **Smoothing** probabilities: \( \mathbb{P}(s_t = i \mid Z^T; \theta) \)
- The normality assumption:
  \[
  f(z_t \mid s_t = i, Z^{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2_{\varepsilon}}} \exp \left\{ \frac{-(z_t - \alpha_0 - \alpha_1 i - \beta_1 z_{t-1} - \cdots - \beta_k z_{t-k})^2}{2\sigma^2_{\varepsilon}} \right\}.
  \]
The equations below form a recursive system:

- The conditional densities of $z_t$ given $Z^{t-1}$ are

$$f(z_t \mid Z^{t-1}; \theta) = \mathbb{I}P(s_t = 0 \mid Z^{t-1}; \theta) f(z_t \mid s_t = 0, Z^{t-1}; \theta) + \mathbb{I}P(s_t = 1 \mid Z^{t-1}; \theta) f(z_t \mid s_t = 1, Z^{t-1}; \theta).$$

- The filtering probabilities of $s_t$ are

$$\mathbb{I}P(s_t = i \mid Z^t; \theta) = \frac{\mathbb{I}P(s_t = i \mid Z^{t-1}; \theta) f(z_t \mid s_t = i, Z^{t-1}; \theta)}{f(z_t \mid Z^{t-1}; \theta)}.$$

- The prediction probabilities are

$$\mathbb{I}P(s_{t+1} = i \mid Z^t; \theta)$$

$$= \mathbb{I}P(s_t = 0, s_{t+1} = i \mid Z^t; \theta) + \mathbb{I}P(s_t = 1, s_{t+1} = i \mid Z^t; \theta)$$

$$= p_{0i} \mathbb{I}P(s_t = 0 \mid Z^t; \theta) + p_{1i} \mathbb{I}P(s_t = 1 \mid Z^t; \theta).$$
Side product: The quasi-log-likelihood function is

$$\mathcal{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ln f(z_t \mid Z^{t-1}; \theta),$$

from which we can solve for the QMLE $\tilde{\theta}_T$.

The estimated filtering and smoothing probabilities are calculated by plugging $\tilde{\theta}_T$ into their formulae.

The expected duration of the $i$th state ($i = 0, 1$) is

$$\sum_{k=1}^{\infty} k p_{ii}^{k-1}(1 - p_{ii}) = 1/(1 - p_{ii});$$

see Hamilton (1989, p. 374). The larger the value of $p_{ii}$, the longer is the expected duration of (the more persistent is) the $i$th state.
Computing Smoothing Probabilities

To compute the smoothing probabilities $\mathbb{P}(s_t = i \mid \mathcal{Z}^T; \theta)$, we adopt the approximation of Kim (1994):

$$\mathbb{P}(s_t = i \mid s_{t+1} = j, \mathcal{Z}^T; \theta) \approx \mathbb{P}(s_t = i \mid s_{t+1} = j, \mathcal{Z}^t; \theta)$$

$$= \frac{\mathbb{P}(s_t = i, s_{t+1} = j \mid \mathcal{Z}^t; \theta)}{\mathbb{P}(s_{t+1} = j \mid \mathcal{Z}^t; \theta)}$$

$$= \frac{p_{ij} \mathbb{P}(s_t = i \mid \mathcal{Z}^t; \theta)}{\mathbb{P}(s_{t+1} = j \mid \mathcal{Z}^t; \theta)},$$

for $i, j = 0, 1$. 
The smoothing probabilities are thus

\[ P(s_t = i \mid Z^T; \theta) = P(s_{t+1} = 0 \mid Z^T; \theta) P(s_t = i \mid s_{t+1} = 0, Z^T; \theta) + P(s_{t+1} = 1 \mid Z^T; \theta) P(s_t = i \mid s_{t+1} = 1, Z^T; \theta) \]

\[ \approx P(s_t = i \mid Z^t; \theta) \times \left( \frac{p_{i0} P(s_{t+1} = 0 \mid Z^T; \theta)}{P(s_{t+1} = 0 \mid Z^t; \theta)} + \frac{p_{i1} P(s_{t+1} = 1 \mid Z^T; \theta)}{P(s_{t+1} = 1 \mid Z^t; \theta)} \right). \]

Using the filtering probability \( P(s_T = i \mid Z^T; \theta) \) as the initial value, we can iterate backward the equations for filtering and prediction probabilities and the equation above to get the smoothing probabilities for \( t = T - 1, \cdots, k + 1 \).
An alternative estimation method is Gibbs sampling which is a Markov Chain Monte Carlo simulation method. This method is Bayesian and treats parameters as random variables.

- Classify $\theta$ into $k$ groups: $\theta = (\theta'_1, \theta'_2, \ldots, \theta'_k)'$.
- By specifying the prior distributions of parameters and likelihood functions, we can derive the conditional posterior distributions:

$$
\pi(\theta_i \mid Z^T, \{\theta_j, j \neq i\}), \quad i = 1, \ldots, k,
$$

which is also known as the full conditional distribution of $\theta_i$.
- Draw parameters from this conditional distribution.
With random initial values \( \theta^{(0)} = (\theta_1^{(0)\prime}, \theta_2^{(0)\prime}, \ldots, \theta_k^{(0)\prime})' \), the recursion for the \( i \)th realization of \( \theta \) proceed as follows.

- Randomly draw a realization \( \theta_1^{(i)} \) from
  \[
  \pi(\theta_1 \mid Z^T, \theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)}).
  \]

- Randomly draw a realization \( \theta_2^{(i)} \) from
  \[
  \pi(\theta_2 \mid Z^T, \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_k^{(i-1)}).
  \]

- Proceeds similarly to draw \( \theta_3^{(i)}, \ldots, \theta_k^{(i)} \) and obtain
  \[
  \theta^{(i)} = (\theta_1^{(i)\prime}, \theta_2^{(i)\prime}, \ldots, \theta_k^{(i)\prime})'.
  \]

- Repeating the procedure above \( N \) times yields the Gibbs sequence:
  \[
  \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}\}.
  \]
The Gibbs sequence converges in distribution exponentially fast to the true distribution of $\theta$, i.e.,

$$\theta^{(N)} \overset{D}{\rightarrow} \pi(\theta \mid Z^T),$$

as $N$ tends to infinity.

For any measurable function $g$,

$$\frac{1}{N} \sum_{i=1}^{N} g(\theta^{(i)}) \overset{\text{a.s.}}{\rightarrow} \mathbb{E}[g(\theta)],$$

where $\overset{\text{a.s.}}{\rightarrow}$ denotes almost sure convergence.
In addition to $\theta$, the unobserved state variables $s_t$, $t = 1, \ldots, T$, are also treated as parameters. The augmented parameter vector is classified into 4 groups:

1. $s_t$, $t = 1, \ldots, T$,
2. $p_{00}$ and $p_{11}$,
3. $\alpha_0$, $\alpha_1$ and $\beta_1, \ldots, \beta_k$,
4. $\sigma^2_{\varepsilon}$.

Random drawings from the conditional posterior distributions yield the Gibbs sequence. To alleviate the effect of initial values, a large number of parameter values in the Gibbs sequence will be discarded.

The sample average of the remaining Gibbs sequence is the desired estimate of unknown parameters.
The null hypothesis is $\alpha_1 = 0$.

- Under the null, the Markov switching model reduces to an AR($k$) model, and the likelihood value is not affected by $p_{00}$ and $p_{11}$. That is, $p_{00}$ and $p_{11}$ are not identified under the null, and they are nuisance parameters.

- When there are unidentified nuisance parameters under the null, the standard likelihood-based tests are invalid, Davies (1977, 1987) and Hansen (1992).
Hansen (1992, 1996) Test

Write $\theta = (\gamma, \theta_1')' = (\alpha_1, p, \theta_1')'$.

- Fixing $\gamma$, the concentrated QMLE of $\theta_1$ is
  $$\hat{\theta}_1(\gamma) = \arg\max_{\theta_1} L_T(\gamma, \theta_1) \overset{P}{\longrightarrow} \theta_1(\gamma).$$

- The concentrated quasi-log-likelihood functions are
  $$\hat{L}_T(\gamma) = L_T(\gamma, \hat{\theta}_1(\gamma)), \quad L_T(\gamma) = L_T(\gamma, \theta_1(\gamma)).$$

- For a given $\gamma$, the likelihood ratio statistics are
  $$\hat{L}_{R_T}(\gamma) = \hat{L}_T(\gamma) - \hat{L}_T(0, p),$$
  $$L_{R_T}(\gamma) = L_T(\gamma) - L_T(0, p).$$
As \( \gamma \) contains nuisance parameters, it is natural to consider the likelihood ratios for all possible values of \( \gamma \). This leads to the supremum statistic:

\[
\sup_{\gamma} \sqrt{T} \hat{\mathcal{L}}\mathcal{R}_T(\gamma).
\]

- Under the null hypothesis,

\[
\sqrt{T} \hat{\mathcal{L}}\mathcal{R}_T(\gamma) = \sqrt{T} [\mathcal{L}R_T(\gamma) - M_T(\gamma)] + \sqrt{T} M_T(\gamma) + o_P(1),
\]

where \( M_T(\gamma) = \mathbb{E}[\mathcal{L}R_T(\gamma)] < 0 \) because \( L_T(\gamma) < L_T(0, p) \) when the null is true (\( \alpha_1 = 0 \)).

- For any \( \gamma \),

\[
\sqrt{T} \hat{\mathcal{L}}\mathcal{R}_T(\gamma) \leq \sqrt{T} Q_T(\gamma) + o_P(1),
\]

where \( Q_T(\gamma) = \mathcal{L}R_T(\gamma) - M_T(\gamma) \). It follows that

\[
\sup_{\gamma} \sqrt{T} \hat{\mathcal{L}}\mathcal{R}_T(\gamma) \leq \sup_{\gamma} \sqrt{T} Q_T(\gamma) + o_P(1).
\]
An empirical-process central limit theorem ensures

$$\sqrt{T} Q_T(\gamma) \Rightarrow Q(\gamma),$$

where $Q$ is a Gaussian process with mean zero and the covariance function $K(\gamma_1, \gamma_2)$. By the continuous mapping theorem,

$$\sup_{\gamma} \sqrt{T} Q_T(\gamma) \overset{\mathbb{P}}{\longrightarrow} \sup_{\gamma} Q(\gamma).$$

sup $Q$ is an upper bound of the supremum statistic:

$$\sup_{\gamma} \sqrt{T} \hat{\mathcal{L}} \mathcal{R}_T(\gamma) \leq \sup_{\gamma} Q(\gamma) + o_\mathbb{P}(1),$$

so that

$$\mathbb{P} \left\{ \sup_{\gamma} \sqrt{T} \hat{\mathcal{L}} \mathcal{R}_T(\gamma) > c \right\} \leq \mathbb{P} \left\{ \sup_{\gamma} Q(\gamma) > c \right\}.$$
• We can simulate $\sup_\gamma Q(\gamma)$ and find its critical values.
  • For a given level, this critical value must be larger than that of $\sup_\gamma \sqrt{T} \hat{\mathcal{LR}}_T(\gamma)$, and this test thus rejects less often than it should.
  • Simulating $Q$ is difficult because we must consider all possible values of $\gamma$. In our application, $\alpha_1$ can take any value on the real line, and $p_{00}$ and $p_{11}$ take any value in $[0, 1]$. Computation depends on the grid points we choose.

• In Hansen (1992, 1996), a standardized supremum statistic is considered:

$$\sup_\gamma \hat{\mathcal{LR}}^*_T(\gamma) = \sup_\gamma \sqrt{T} \hat{\mathcal{LR}}_T(\gamma)/\hat{\mathcal{V}}_T(\gamma)^{1/2},$$

where $\hat{\mathcal{V}}_T(\gamma)$ is a variance estimate.
To test independence of state variables, the null hypotheses are

\[ p_{00} = p_{10}, \quad \text{and} \quad p_{01} = p_{11}. \]

The null hypotheses can be expressed as

\[ p_{00} + p_{11} = 1, \]

which can be tested using standard likelihood-based tests, such as the Wald test.

Other linear (or nonlinear) hypotheses can also be tested using standard likelihood-based tests.
Hsu and Kuan (2001): Apply a bivariate Markov switching model to Taiwan’s real GDP and employment growth rates and estimate it via Gibbs sampling.

Business cycles:

- Lucas (1977): Comovement of important macroeconomic variables such as production, consumption, investment and employment.
- Diebold and Rudebusch (1996): A model for business cycles should take into account the comovement of economic variables and persistence of economic states.
- Blanchard and Quah (1989): Analyzing GDP alone is not enough to characterize the effects of both supply and demand shocks.
Let $\zeta_t$ denote the vector of GDP and employment. Taking seasonal differences of $\ln(\zeta_t)$ yields the annual growth rates of $\zeta_t$:

$$z_t = \ln(\zeta_t) - \ln(\zeta_{t-4}).$$

For the full sample (1979 Q1 – 1999 Q3), the smoothing probabilities $\Pr(s_t = 1 \mid Z^T)$ indicate that these probabilities are almost zero in 1990s and hence do not identify any cycles.

The maximal-Wald test of Andrews (1993) rejects the null hypothesis of no mean change in the full sample at 5% level.

The least-squares change-point estimates further indicate that the change point for the GDP growth rates was 1989 Q4 and that for the employment growth rates was 1987 Q4. We thus also focus on the after-change sample of $z_t$ from 1989 Q4 through 1999 Q3.
Figure: The growth rates of GDP (left) and employment (right): 1979 Q1–1999 Q3

Note: The average growth rates of GDP and employment are 7.81% resp. 2.56% before 1990 and drop to 6.19% resp. 1.28% after 1990.
Figure: The smoothing prob. of $s_t = 1$: bivariate model, 1979 Q1–1999 Q3
Figure: The smoothing prob. of $s_t = 1$: bivariate model, 1990 Q1–1999 Q3
Estimated average growth rates of GDP: 7.35% vs. 3.26% for after-change sample.

- Huang (1999): 11.3% vs. 7.3%
- Huang, Kuan and Lin (1998): 10.12% vs. 5.74%

Estimated average growth rates of employment: 1.46% vs. 1.15%

Estimated durations: 3.2 vs. 2.3 quarters

- Huang (1999): 5 vs. 13.7 quarters
- Huang, Kuan and Lin (1998): 22.7 vs. 13.7 quarters

Peaks and troughs: determined by the smoothing probabilities with 0.5 as the cut-off value

- This study: (1995 Q2 and 1995 Q4), (1997 Q4 and 1998 Q4)
Figure: The smoothing prob. of $s_t = 1$: univariate model for GDP (left) and employment (right), 1990 Q1–1999 Q3
GARCH \((p, q)\) model: \(z_t = \sqrt{h_t} \varepsilon_t\), with

\[
h_t = c + \sum_{i=1}^{q} a_i z_{t-i}^2 + \sum_{i=1}^{p} b_i h_{t-i},
\]

the conditional variance of \(z_t\) given the information up to time \(t - 1\).

- **GARCH(1,1):**

  \[
h_t = c + a_1 z_{t-1}^2 + b_1 h_{t-1}.
\]

  It is an IGARCH if \(a_1 + b_1 = 1\).

- Lamoureux and Lastrapes (1990): The detected IGARCH pattern may be a consequence of ignored parameter changes in the model.
Switching ARCH Models

- Switching ARCH of Cai (1994): \( z_t = \sqrt{h_t} \varepsilon_t \), and
  \[
  h_t = \alpha_0 + \alpha_1 s_t + \sum_{i=1}^{q} a_i z_{t-i}^2.
  \]

- Switching ARCH of Hamilton and Susmel (1994): \( z_t = \sqrt{\lambda_s} \zeta_t \), \( \zeta_t = \sqrt{\eta_t} \varepsilon_t \) and
  \[
  \eta_t = c + \sum_{i=1}^{q} a_i \zeta_{t-i}^2.
  \]

The conditional variances in two regimes are proportional to each other:

\[
\text{var}(z_t \mid s_t = i, \Phi_{t-1}) = \lambda_i \eta_t, \quad i = 0, 1.
\]
Can we consider a switching GARCH model, such as

\[ h_t = \alpha_0 + \alpha_1 s_t + a_1 z^2_{t-1} + b_1 h_{t-1}? \]

- If the conditional variance \( h_t \) depends on \( h_{t-1} \), then \( h_t \) depends not only on \( s_t \) but also on \( s_{t-1} \). The dependence of \( h_{t-1} \) on \( h_{t-2} \) then implies that \( h_t \) is also affected by the value of \( s_{t-2} \), and so on. That is, \( h_t \) is path dependent.

- The conditional variance at time \( t \) is determined by \( 2^t \) possible realizations of \((s_t, s_{t-1}, \ldots, s_1)\). Model becomes very complex and estimation is intractable.
Gray (1996): \( z_t = \sqrt{h_{i,t}} \varepsilon_t \), where \( h_{i,t} = \text{var}(z_t \mid s_t = i, \Phi_{t-1}) \) is a GARCH\((p, q)\) process:

\[
    h_{i,t} = c_i + \sum_{j=1}^{q} a_{i,j} z_{t-j}^2 + \sum_{j=1}^{p} b_{i,j} h_{t-j}.
\]

Gray suggests computing \( h_t \) as weighted sums of \( h_{i,t} \) with the weights being the prediction probabilities \( \text{IP}(s_t = i \mid \Phi_{t-1}) \):

\[
    h_t = \text{IE}(z_t^2 \mid \Phi_{t-1}) = h_{0,t} \text{IP}(s_t = 0 \mid \Phi_{t-1}) + h_{1,t} \text{IP}(s_t = 1 \mid \Phi_{t-1}).
\]

There is no need to consider all possible values of \( (s_t, \ldots, s_1) \).
Following Gray (1996), it is now easy to construct a model with switching conditional mean and variance. For example, $z_t = \mu_{i,t} + v_{i,t}$, $i = 0, 1$, where

$$\mu_{i,t} = \mathbb{E}(z_t \mid s_t = i, \Phi_{t-1}),$$

$$v_{i,t} = \sqrt{h_{i,t}} \varepsilon_t,$$

and

$$h_{i,t} = c_i + \sum_{j=1}^{q} \ a_{i,j} v_{t-j}^2 + \sum_{j=1}^{p} \ b_{i,j} h_{t-j}.$$
The conditional mean and variance are

\[ h_t = \mathbb{E}(z_t^2 \mid \Phi_{t-1}) - \mathbb{E}(z_t \mid \Phi_{t-1})^2, \]

\[ v_t = z_t - \mathbb{E}(z_t \mid \Phi_{t-1}), \] where

\[ \mathbb{E}(z_t \mid \Phi_{t-1}) = \mu_{0,t} \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) + \mu_{1,t} \mathbb{P}(s_t = 1 \mid \Phi_{t-1}), \]

\[ \mathbb{E}(z_t^2 \mid \Phi_{t-1}) = \mathbb{E}(z_t^2 \mid s_t = 0, \Phi_{t-1}) \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) \]

\[ + \mathbb{E}(z_t^2 \mid s_t = 1, \Phi_{t-1}) \mathbb{P}(s_t = 1 \mid \Phi_{t-1}) \]

\[ = (\mu_{0,t}^2 + h_{0,t}) \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) \]

\[ + (\mu_{1,t}^2 + h_{1,t}) \mathbb{P}(s_t = 1 \mid \Phi_{t-1}). \]
A leading model of $\Delta r_t$ is

$$\Delta r_t = \alpha_0 + \beta_0 r_{t-1} + \nu_t,$$

where $\nu_t = \sqrt{h_t \epsilon_t}$ with $h_t = c_0 + a_0 \nu_{t-1}^2 + b_0 h_{t-1}$; see e.g., Chan et al. (1992). Letting $\mu$ denote the long-run level of $r_t$, $\alpha_0 = \rho \mu$ and $\beta_0 = -\rho$, the model above becomes

$$\Delta r_t = \rho (\mu - r_{t-1}) + \nu_t.$$

As long as $\rho > 0$ (i.e., $\beta_0 < 0$), $\Delta r_t$ is positive (negative) when $r_{t-1}$ is below (above) the long-run level. In this case, $r_t$ will adjust toward the long-run level and hence exhibit mean reversion.
Following Gray (1996), we postulate

\[ \Delta r_t = \alpha_i + \beta_i r_{t-1} + v_{i,t}, \quad i = 0, 1, \]

and \[ v_{i,t} = \sqrt{h_{i,t}} \varepsilon_t \] with

\[ h_{i,t} = c_i + a_i v_{t-1}^2 + b_i h_{t-1}, \quad i = 0, 1. \]

The data are the weekly average rates of the 30-day Commercial Paper in the money market, from Jan. 4, 1994 through Dec. 7, 1998.
Figure: The weekly interest rates $r_t$: Jan. 1994–Dec. 1998.
Figure: The estimated smoothing probabilities of \( s_t = 0 \).
Figure: The estimated conditional variances $h_t$. 
Concluding Remarks

- There are many potential empirical applications.
- A proper and computationally simpler test for switching parameters is badly needed.
- Construct general testing procedure when nuisance parameters are not identified under the null.
- The Markovian switching mechanism may also be imposed on other models to yield new models and different applications.