LECTURE ON
THE MARKOV SWITCHING MODEL

CHUNG-MING KUAN

Department of Finance & CRETA

National Taiwan University

April 10, 2011

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. Extension: Innovation Regime Switching Model
   - IRS Model
   - Comparison with Other Models
   - Dynamic Properties
   - Empirical Study

Time Series Models

- Linear models for conditional mean: AR, MA, ARMA, and ARMAX
- Nonlinear Models for conditional mean: NLAR, AR with random coefficients, threshold models, Markov switching model, artificial neural networks; Tong (1990) and Granger and Teräsvirta (1993)
- 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

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 $\mathbf{1}_{\{\lambda_t \le c\}}$

MS Model of Conditional Mean

Hamilton (1989, *Econometrica*): Let s_t be an unobservable state variable governed by a first order Markov chain with the transition matrix:

$$\mathbf{P} = \begin{bmatrix} \mathbb{P}(s_t = 0 \mid s_{t-1} = 0) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 0) \\ \mathbb{P}(s_t = 0 \mid s_{t-1} = 1) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 1) \end{bmatrix}$$

$$= \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix},$$

so that z_t are jointly determined by ε_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.

Some 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:

$$\mathbf{z}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 s_t + \mathbf{B}_1 \mathbf{z}_{t-1} + \cdots + \mathbf{B}_k \mathbf{z}_{t-k} + \boldsymbol{\varepsilon}_t.$$

- 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

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

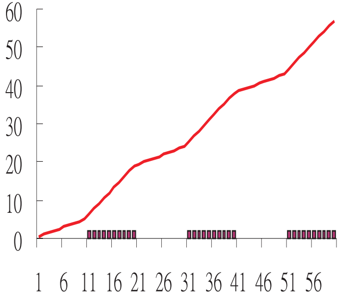


Figure: The Markov trend function with $\alpha_1 > 0$ (left) and $\alpha_1 < 0$ (right).

Quasi-Maximum Likelihood Estimation

- The model parameters: $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_k, \sigma_\varepsilon^2, p_{00}, p_{11})'$.
- Optimal forecasts of $s_t = i$ $(i = 0, 1)$ based on different information sets:
  - Prediction probabilities: $\mathbb{P}(s_t = i \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$, with $\mathcal{Z}^{t-1} = \{z_{t-1}, \ldots, z_1\}$
  - Filtering probabilities: $\mathbb{P}(s_t = i \mid \mathcal{Z}^{t}; \boldsymbol{\theta})$
  - Smoothing probabilities: $\mathbb{P}(s_t = i \mid \mathcal{Z}^{T}; \boldsymbol{\theta})$
- The normality assumption:

$$f(z_t \mid s_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ \frac{-(z_t - \alpha_0 - \alpha_1\, i - \beta_1 z_{t-1} - \cdots - \beta_k z_{t-k})^2}{2\sigma_\varepsilon^2} \right\}.$$

The equations below form a recursive system:

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

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

- The filtering probabilities of s_t are

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

- The prediction probabilities are

$$\begin{aligned} &\mathbb{P}(s_{t+1} = i \mid \mathcal{Z}^{t}; \boldsymbol{\theta}) \\ &\quad = \mathbb{P}(s_t = 0, s_{t+1} = i \mid \mathcal{Z}^{t}; \boldsymbol{\theta}) + \mathbb{P}(s_t = 1, s_{t+1} = i \mid \mathcal{Z}^{t}; \boldsymbol{\theta}) \\ &\quad = p_{0i}\, \mathbb{P}(s_t = 0 \mid \mathcal{Z}^{t}; \boldsymbol{\theta}) + p_{1i}\, \mathbb{P}(s_t = 1 \mid \mathcal{Z}^{t}; \boldsymbol{\theta}). \end{aligned}$$

- Side product: The quasi-log-likelihood function is

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

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

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

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

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

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

$$\begin{aligned}
&\mathbb{P}(s_t = i \mid s_{t+1} = j, \mathcal{Z}^T; \boldsymbol{\theta}) \\
&\qquad \approx \mathbb{P}(s_t = i \mid s_{t+1} = j, \mathcal{Z}^t; \boldsymbol{\theta}) \\
&\qquad = \frac{\mathbb{P}(s_t = i, s_{t+1} = j \mid \mathcal{Z}^t; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = j \mid \mathcal{Z}^t; \boldsymbol{\theta})} \\
&\qquad = \frac{p_{ij}\, \mathbb{P}(s_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = j \mid \mathcal{Z}^t; \boldsymbol{\theta})},
\end{aligned}$$

for $i, j = 0, 1$.

The smoothing probabilities are thus

$$
\begin{aligned}
&\mathbb{P}(s_t = i \mid \mathcal{Z}^T; \boldsymbol{\theta}) \\
&\quad = \mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta})\ \mathbb{P}(s_t = i \mid s_{t+1} = 0, \mathcal{Z}^T; \boldsymbol{\theta}) \\
&\qquad\qquad + \mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta})\ \mathbb{P}(s_t = i \mid s_{t+1} = 1, \mathcal{Z}^T; \boldsymbol{\theta}) \\
&\quad \approx \mathbb{P}(s_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta}) \\
&\qquad\qquad \times \left(\frac{p_{i0}\ \mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta})} + \frac{p_{i1}\ \mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta})} \right).
\end{aligned}
$$

Using the filtering probability $\mathbb{P}(s_T = i \mid \mathcal{Z}^T; \boldsymbol{\theta})$ 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 $t = T-1, \cdots, k+1$.

Estimation via Gibbs Sampling

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

  $$\pi\big(\boldsymbol{\theta}_i \mid \mathcal{Z}^T, \{\boldsymbol{\theta}_j, j \neq i\}\big), \qquad i = 1, \ldots, k,$$

  which is also known as the full conditional distribution of $\boldsymbol{\theta}_i$.
- Draw parameters from this conditional distribution.

- With random initial values $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)\prime}, \boldsymbol{\theta}_2^{(0)\prime}, \ldots, \boldsymbol{\theta}_k^{(0)\prime})'$, the recursion for the i th realization of $\boldsymbol{\theta}$ proceed as follows.
    - Randomly draw a realization $\boldsymbol{\theta}_1^{(i)}$ from
$$\pi\big(\boldsymbol{\theta}_1 \mid \mathcal{Z}^T, \boldsymbol{\theta}_2^{(i-1)}, \ldots, \boldsymbol{\theta}_k^{(i-1)}\big).$$
    - Randomly draw a realization $\boldsymbol{\theta}_2^{(i)}$ from
$$\pi\big(\boldsymbol{\theta}_2 \mid \mathcal{Z}^T, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_3^{(i-1)}, \ldots, \boldsymbol{\theta}_k^{(i-1)}\big).$$
    - Proceeds similarly to draw $\boldsymbol{\theta}_3^{(i)}, \ldots, \boldsymbol{\theta}_k^{(i)}$ and obtain
$$\boldsymbol{\theta}^{(i)} = (\boldsymbol{\theta}_1^{(i)\prime}, \boldsymbol{\theta}_2^{(i)\prime}, \ldots, \boldsymbol{\theta}_k^{(i)\prime})'.$$
    - Repeating the procedure above N times yields the Gibbs sequence:
$$\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(N)}\}.$$

Geman and Geman (1984)

- The Gibbs sequence converges in distribution exponentially fast to the true distribution of $\boldsymbol{\theta}$, i.e.,

$$\boldsymbol{\theta}^{(N)} \xrightarrow{D} \pi\big(\boldsymbol{\theta} \mid \mathcal{Z}^T\big),$$

as N tends to infinity.

- For any measurable function g,

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

where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence.

A Summary

- In addition to $\boldsymbol{\theta}$, 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. α_0, α_1 and $\beta_1, \ldots, \beta_k$,
    4. σ_ε^2.
- 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.

Testing for Switching 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 $\boldsymbol{\theta} = (\gamma, \boldsymbol{\theta}_1')' = (\alpha_1, \mathbf{p}, \boldsymbol{\theta}_1')'$.

- Fixing γ, the concentrated QMLE of $\boldsymbol{\theta}_1$ is

$$\hat{\boldsymbol{\theta}}_1(\gamma) = \operatorname{argmax} L_T(\gamma, \boldsymbol{\theta}_1) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_1(\gamma).$$

- The concentrated quasi-log-likelihood functions are

$$\hat{L}_T(\gamma) = L_T\big(\gamma, \hat{\boldsymbol{\theta}}_1(\gamma)\big), \quad L_T(\gamma) = L_T(\gamma, \boldsymbol{\theta}_1(\gamma)).$$

- For a given γ, the likelihood ratio statistics are

$$\widehat{\mathcal{LR}}_T(\gamma) = \hat{L}_T(\gamma) - \hat{L}_T(0, \mathbf{p}),$$

$$\mathcal{LR}_T(\gamma) = L_T(\gamma) - L_T(0, \mathbf{p}).$$

As γ contains nuisance parameters, it is natural to consider the likelihood ratios for all possible values of γ. This leads to the supremum statistic: $\sup_\gamma \sqrt{T}\,\widehat{\mathcal{LR}}_T(\gamma)$.

- Under the null hypothesis,

$$\sqrt{T}\,\widehat{\mathcal{LR}}_T(\gamma) = \sqrt{T}\,[\mathcal{LR}_T(\gamma) - M_T(\gamma)] + \sqrt{T}\,M_T(\gamma) + o_{\mathbb{P}}(1),$$

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

- For any γ,

$$\sqrt{T}\,\widehat{\mathcal{LR}}_T(\gamma) \leq \sqrt{T}\,Q_T(\gamma) + o_{\mathbb{P}}(1),$$

  where $Q_T(\gamma) = \mathcal{LR}_T(\gamma) - M_T(\gamma)$. It follows that

$$\sup_\gamma \sqrt{T}\,\widehat{\mathcal{LR}}_T(\gamma) \leq \sup_\gamma \sqrt{T}\,Q_T(\gamma) + o_{\mathbb{P}}(1).$$

- 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,

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

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

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

so that

$$\mathbb{P}\left\{ \sup_{\gamma} \sqrt{T} \widehat{\mathcal{LR}}_T(\gamma) > c \right) \leq \mathbb{P}\left\{ \sup_{\gamma} Q(\gamma) > c \right\}.$$

- 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 $\sup_\gamma \sqrt{T}\,\widehat{\mathcal{LR}}_T(\gamma)$, and this test thus rejects less often than it should.
  - Simulating Q is difficult because we must consider all possible values of γ. In our application, α_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:

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

where $\hat{V}_T(\gamma)$ is a variance estimate.

Testing Other Hypotheses

- 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.

Application: Taiwan's Business Cycles

- 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 ζ_t denote the vector of GDP and employment. Taking seasonal differences of $\ln(\zeta_t)$ yields the annual growth rates of ζ_t:

$$\mathbf{z}_t = \ln(\zeta_t) - \ln(\zeta_{t-4}).$$

- For the full sample (1979 Q1 – 1999 Q3), the smoothing probabilities $\mathbb{P}(s_t = 1 \mid \mathcal{Z}^T)$ 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 the after-change sample of $\mathbf{z}_t$ from 1989 Q4 through 1999 Q3.

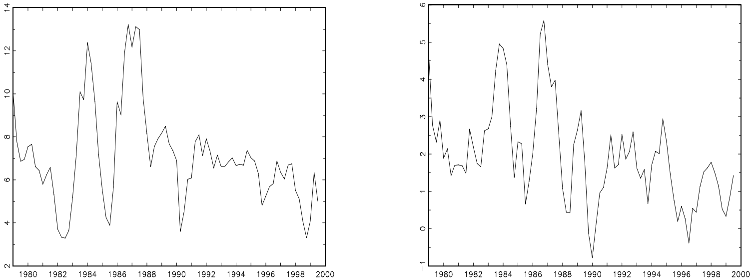


Figure: The growth rates of GDP (left) and employment (right): 1979 Q1–1999 Q3

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.

Bivariate MS Result: Full Sample

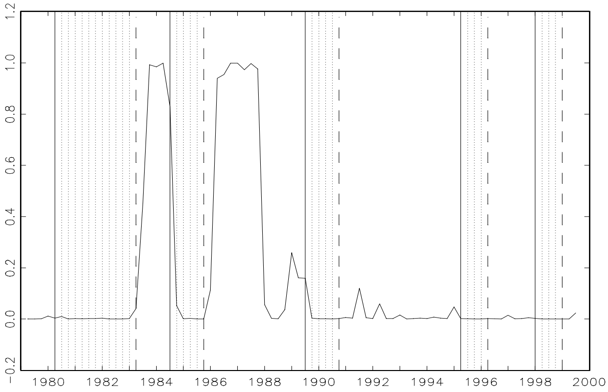


Figure: The smoothing prob. of $s_t = 1$: bivariate model, 1979 Q1–1999 Q3

Bivariate MS Result: After-Change Sample

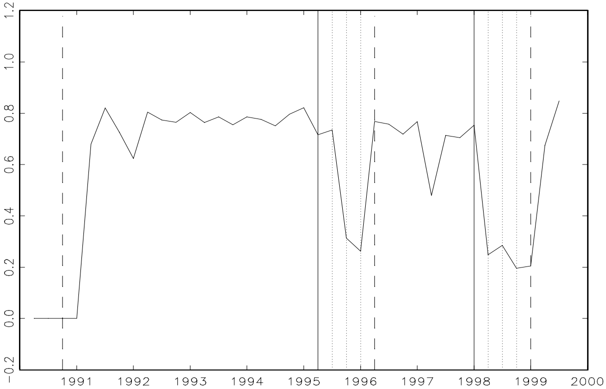


Figure: The smoothing prob. of $s_t = 1$: bivariate model, 1990 Q1–1999 Q3

Estimation Results

- 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 quareter
  - 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)
  - CEPD: (1995 Q1 and 1996 Q1), (1997 Q4 and 1998 Q4).

Univariate MS Result: After-Change Sample

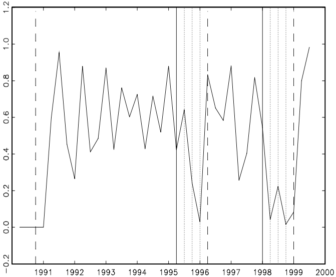


Figure: The smoothing prob. of $s_t = 1$: univariate model for GDP (left) and employment (right), 1990 Q1–1999 Q3

MS Model of Conditional Variance

- 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): $z_t = \sqrt{\lambda_{s_t}}\,\zeta_t$, $\zeta_t = \sqrt{\eta_t}\,\varepsilon_t$ 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:

$$\mathrm{var}(z_t \mid s_t = i, \Phi_{t-1}) = \lambda_i \eta_t, \quad i = 0, 1.$$

Switching GARCH Models

Can we consider a switching GARCH model, such as

$$h_t = \alpha_0 + \alpha_1 s_t + a_1 z_{t-1}^2 + b_1 h_{t-1}?$$

- 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. *JFE*): $z_t = \sqrt{h_{i,t}}\,\varepsilon_t$, where $h_{i,t} = \operatorname{var}(z_t \mid s_t = i, \Phi_{t-1})$ 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 $\mathbb{P}(s_t = i \mid \Phi_{t-1})$:

$$h_t = \mathbb{E}(z_t^2 \mid \Phi_{t-1}) = h_{0,t}\,\mathbb{P}(s_t = 0 \mid \Phi_{t-1}) + h_{1,t}\,\mathbb{P}(s_t = 1 \mid \Phi_{t-1}).$$

There is no need to consider all possible values of $(s_t, \ldots, s_1)$.

MS Model of Conditional Mean and Variance

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

$$\begin{aligned}
\mathbb{E}(z_t \mid \Phi_{t-1}) &= \mu_{0,t}\, \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) + \mu_{1,t}\, \mathbb{P}(s_t = 1 \mid \Phi_{t-1}), \\
\mathbb{E}(z_t^2 \mid \Phi_{t-1}) &= \mathbb{E}(z_t^2 \mid s_t = 0, \Phi_{t-1})\, \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) \\
&\qquad + \mathbb{E}(z_t^2 \mid s_t = 1, \Phi_{t-1})\, \mathbb{P}(s_t = 1 \mid \Phi_{t-1}) \\
&= (\mu_{0,t}^2 + h_{0,t})\, \mathbb{P}(s_t = 0 \mid \Phi_{t-1}) \\
&\qquad + (\mu_{1,t}^2 + h_{1,t})\, \mathbb{P}(s_t = 1 \mid \Phi_{t-1}).
\end{aligned}$$

Application: Taiwan's Short Term Interest Rates

A leading model of Δr_t is

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

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

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

As long as $\rho > 0$ (i.e., $\beta_0 < 0$), Δr_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), Lin, Hung, and Kuan (2002) 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.

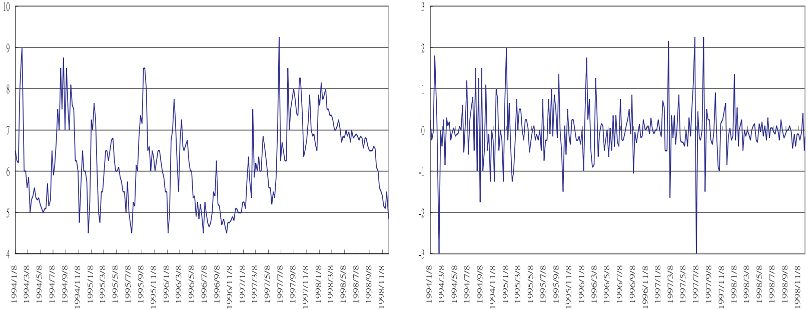


Figure: The weekly interest rates r_t: Jan. 1994–Dec. 1998.

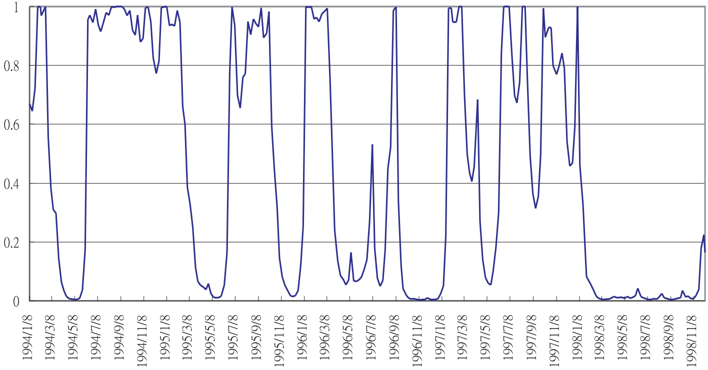


Figure: The estimated smoothing probabilities of $s_t = 0$.

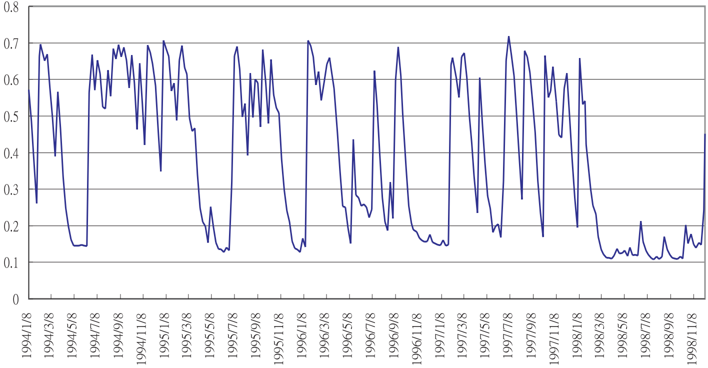


Figure: The estimated conditional variances h_t.

Extension: Innovation Regime Switching Model

- Existing Markov switching models
  - There is only one functional relationship; different regimes are characterized by state-dependent parameters.
  - The dynamic patterns in different regimes are similar.
- Motivations of a new regime switching model
  - Different switching mechanism
  - Regimes are characterized by distinct functions and hence can accommodate different dynamic behaviors.
- Leading patterns in macroeconomic and financial time series.
  - Unit-root nonstationarity: Nelson and Plosser (1982), Campbell and Mankiw (1987)
  - Trend stationarity (with or without breaks): Blanchard (1981), Clark (1987), Perron (1989)

IRS Model

Kuan, Huang, and Tsay (2005, *JBES*): $y_t = y_{1,t} + y_{0,t}$ with

$$y_{1,t} = g(y_{1,t-1}, \ldots, y_{1,t-p}; \boldsymbol{\theta}_1) + s_t \upsilon_t,$$

$$y_{0,t} = h(y_{0,t-1}, \ldots, y_{0,t-q}; \boldsymbol{\theta}_0) + (1 - s_t)\upsilon_t,$$

where $s_t = 0, 1$ is the state variable following a Markov chain.

- Different switching mechanism: s_t is linked to innovations.
- Each innovation excites only one component; e.g., when $s_t = 1$, $y_{1,t}$ is excited , but $y_{0,t}$ evolves without υ_t.
- The concurrent state s_t determines the effect of υ_t on y_{t+j}.
- May exhibit completely different dynamics when $g \neq h$.

Proposed IRS(1; m, n) Model

A specific model: $y_t = y_{1,t} + y_{0,t}$ with

$$\begin{aligned}(1-B)y_{1,t} &= \alpha_0 + s_t \upsilon_t = (\alpha_0 + s_t\alpha_1) + s_t\varepsilon_t,\\ \Psi(B)y_{0,t} &= \Phi(B)(1-s_t)\upsilon_t\\ &= \Phi(B)(1-s_t)\alpha_1 + \Phi(B)(1-s_t)\varepsilon_t.\end{aligned}$$

- It consists of a random walk component (with drift) and a stationary ARMA component, so that the dynamics may be unit-root nonstationarity or covariance stationarity.
- The effect of ε_t on future y_t may be permanent or transitory.
- This model constitutes intermediate cases between a random walk (with drift) and a (trend-)stationary model.

A decomposition:

$$
\begin{aligned}
y_t = \alpha_0 t + \alpha_1 \sum_{i=1}^{t} s_i + \alpha_1 \Psi(B)^{-1}\Phi(B)(1 - s_t) + \sum_{i=1}^{t} s_i \varepsilon_i \\
+ \Psi(B)^{-1}\Phi(B)(1 - s_t)\varepsilon_t.
\end{aligned}
$$

- Deterministic and stochastic trends with Stationary and nonstationary time path
- Trend with or without endogenous breaks (linked to s_t)
- The transition between trend segments is smooth.

Note: Many other variants are possible, e.g., switching between a long-memory component and a covariance stationary component.

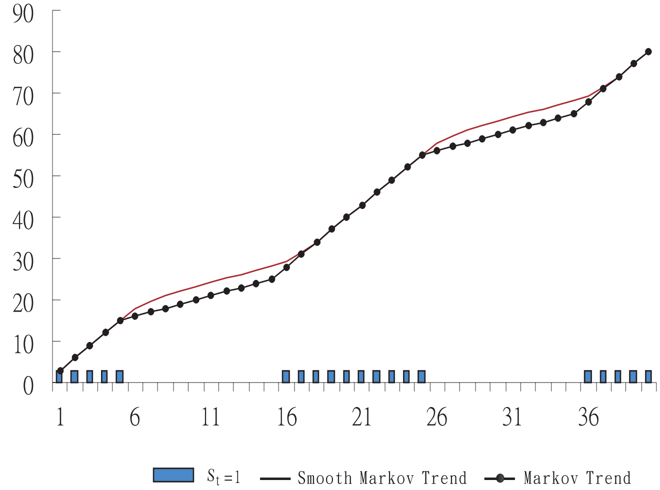


Figure: Simulated Markov trend and smooth Markov trend.

Comparison with Other Models

- Evans and Wachtel (1993): $y_t = s_t y_{1,t} + (1 - s_t) y_{0,t}$ with

$$
\begin{aligned}
y_{1,t} &= y_{1,t-1} + v_t, \\
y_{0,t} &= \psi_1 y_{0,t-1} + u_t, \qquad |\psi_1| < 1.
\end{aligned}
$$

  - This model switches between two processes, so that all all past innovations are affected by switching.
  - The time path exhibits big and sudden jumps.
  - Ironically, switching affects the past but has no effect on what will happen in the future.

- McCabe and Tremayne (1995): $y_t = a_t y_{t-1} + u_t$,

$$y_t = u_t + \sum_{i=1}^{t-1} \left(\prod_{j=0}^{i-1} a_{t-j} \right) u_{t-j}.$$

  Stochastic unit root of Granger and Swanson (1997): $a_t = \exp(\alpha_t)$.

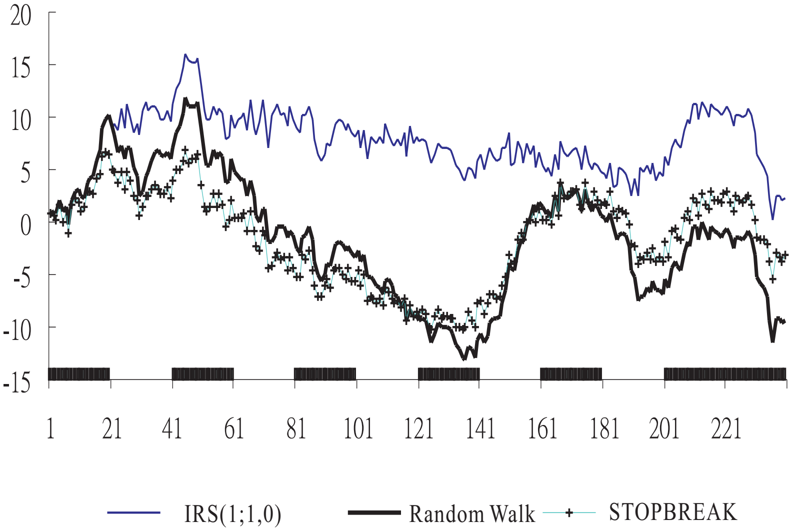
  - It is difficult to interpret when the parameter switches from a stable region to an explosive region.
  - Switching also affects the past but not the future.

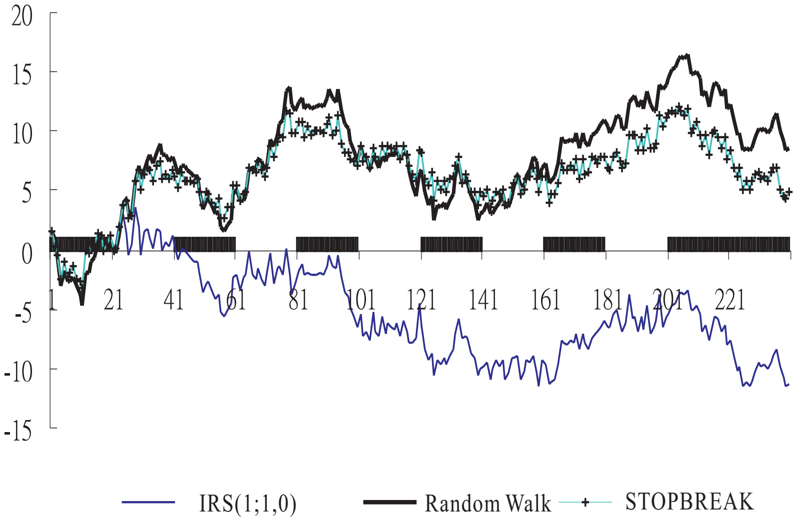
- Engle and Smith (1999): The STOPBREAK process is

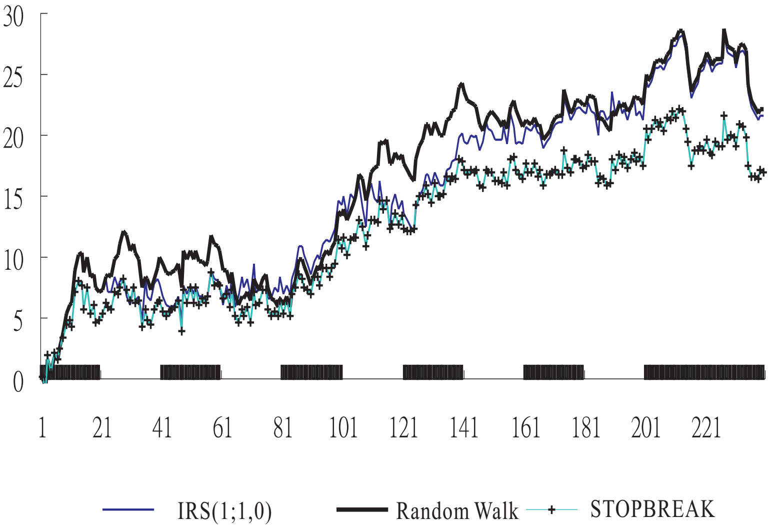
$$y_t = \sum_{i=1}^{\infty} q_{t-i} \upsilon_{t-i} + \upsilon_t,$$

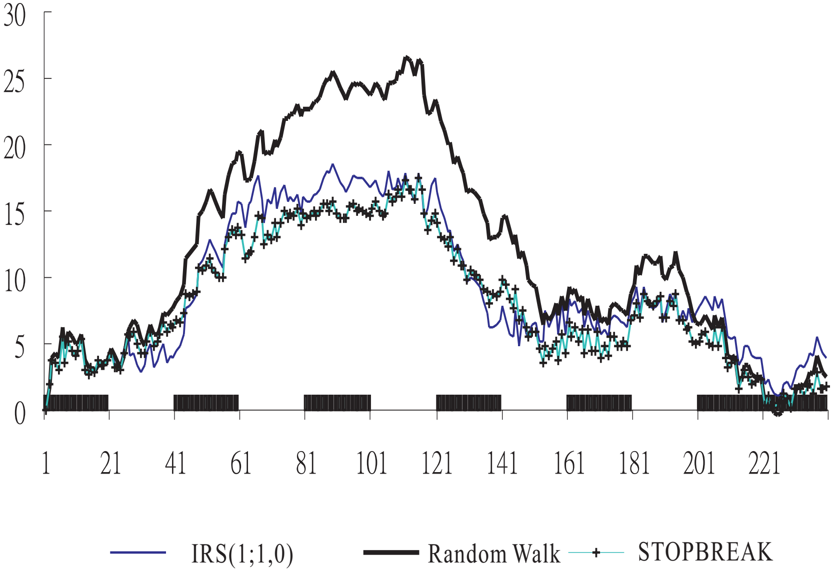
where $q_{t-i} = \upsilon_{t-i}^2 / (\gamma + \upsilon_{t-i}^2)$ with $\gamma \geq 0$.

  - It is very close to an $I(1)$ process $(0 < q_t \approx 1)$.
  - y_t are positively correlated.
  - It does not really exhibit stationary behavior.


20
15
10
5
0
-5
-10
-15
1
21
41
61
81
101
121
141
161
181
201
221
IRS(1;1,0)
Random Walk
STOPBREAK


20
15
10
5
0
-5
-10
-15
1
21
41
61
81
101
121
141
161
181
201
221
IRS(1;1,0)
Random Walk
STOPBREAK


30
25
20
15
10
5
0
1
21
41
61
81
101
121
141
161
181
201
221
IRS(1;1,0)
Random Walk
STOPBREAK


30
25
20
15
10
5
0
1
21
41
61
81
101
121
141
161
181
201
221
IRS(1;1,0)
Random Walk
STOPBREAK

Dynamic Properties

- Assume: $\mathbb{E}(\varepsilon_t|S^t) = 0$, $\text{var}(\varepsilon_t|S^t) = \sigma_\upsilon^2$, $\text{cov}(\varepsilon_t, \varepsilon_{t-i}|S^t) = 0$.
- When $\alpha_1 = 0$, y_t is the sum of uncorrelated components:

$$y_t = \alpha_0 t + \sum_{i=1}^{t} s_i \varepsilon_i + \Psi(B)^{-1}\Phi(B)(1 - s_t)\varepsilon_t.$$

Then, $\mathbb{E}(y_t) = \alpha_0\, t$ and

$$\text{var}(y_t) = \sigma_\varepsilon^2 \sum_{i=1}^{t} \mathbb{P}(s_i = 1) + \sigma_\varepsilon^2 \sum_{i=1}^{t} (\psi_i^*)^2 [1 - \mathbb{P}(s_i = 1)].$$

  - $\text{var}(y_t)$ grows without bound if $\sum_{i=1}^{t} \mathbb{P}(s_i = 1)$ diverges. When $0 < \mathbb{P}(s_i = 1) = \pi_0 < 1$, $\text{var}(y_t)$ is linear in t.
  - $\text{var}(y_t)$ is finite when $\sum_{t=-\infty}^{\infty} \mathbb{P}(s_t = 1) < \infty$. Then, $s_t = 1$ for at most finitely many t with prob. one (Borel-Cantelli).

- When $\alpha_1 \neq 0$, y_t is the sum of correlated components, but $\mathrm{var}(y_t)$ still depends on $\sum_{i=1}^{t} \mathbb{P}(s_i = 1)$.
- The long-run effect of ε_t on the optimal forecast of y_{t+k} is

$$\delta_t \equiv \lim_{k\to\infty} \frac{\partial\, \mathbb{E}(y_{t+k}|\mathcal{F}^t)}{\partial \varepsilon_t} = s_t,$$

  which may be one or zero.
- The IRS model has an ARMA representation with random MA coefficients.
- $z_t = y_t - y_{t-1}$ is covariance stationary (asymptotically).

Empirical Study

- Estimation: QMLE based on the Hamilton filter or the state-space model
- Testing $p_{11} = 1$: There are nuisance parameters not identified under the null; Hansen's test or a simulation-based test
- Empirical Study of U.S. real GDP (1947:I – 2002:I), 221 observations
  - The chosen model is IRS(1;2,2) with $\hat{p}_{11} \approx 0.91$ and $\hat{p}_{00} \approx 0.60$.
  - This is not an ARIMA process: p-value of 0.91 is 0.034.
  - s_t are not independent over time: The Wald test of the null $p_{00} + p_{11} = 1$ (i.e., $p_{01} = p_{11}$) is 18.79, significant at 1% level.

- About 84% of the observations have $\mathbb{P}(s_t = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta}) > 0.5$ and hence are more likely to exhibit unit-root nonstationarity.
  - Shocks are not always permanent, in contrast with the results based on unit-root models.
  - Permanent shocks are more frequent than those predicted by trend-break models.
- Nonstationarity (stationarity) periods match NBER expansions (recessions) very closely; cf. Beaudry and Koop (1993).
  - The estimated growth rates are -0.83% for recessions and 1.15% for expansion (NBER: -0.35% and 1.10%).
  - The expected durations of recession and expansion are 2.6 and 11 quarters (NBER: 3.4 and 18.4 quarters).

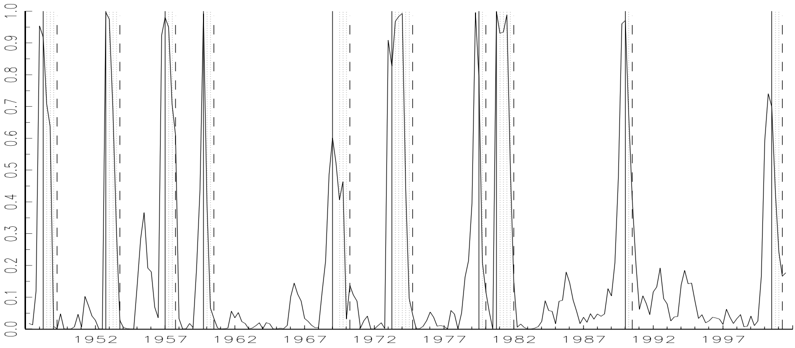


Figure: Estimated smoothing probabilities of $s_t = 0$.

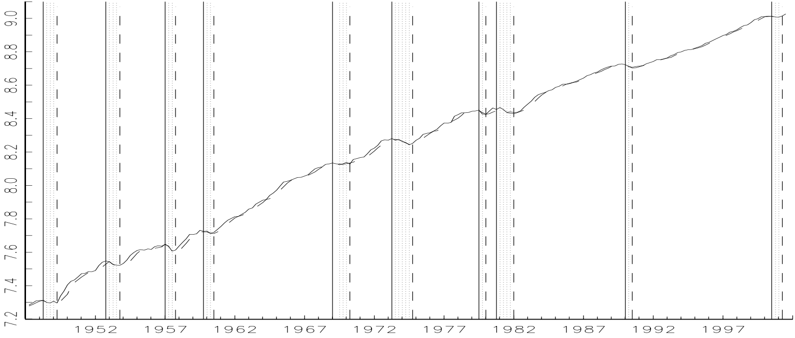


Figure: The expected trend line in U.S. real GDP.