Chapter Five

Regime Switching Models

Tingguo Zheng (zhengtg@gmail.com)
School of Economics & WISE & Chow, Xiamen University

(For Chow & SOE & WISE Msc and PhD Program, 2026-2027, Autumn Semester)

- **Reference:**
  - “Analysis of Financial Time Series”, 2010, Tsay, Chapter 4.
  - “Nonlinear Time Series Models in Empirical Finance”, 2000, Franses and van Dijk, Chap 3.

- **Outline:**
  - Nonlinear Time Series Models: TAR model, STAR model, MS model;
  - Estimation of Nonlinear Models: SETAR model, STAR model, MS model
  - Nonlinearity tests: Threshold tests, Test for STAR nonlinearity, Test for MS-type nonlinearity.

1 Nonlinear Models

1.1 Threshold Autoregressive (TAR) Models-Tong (1978, 1990)

- TAR model is one popular class of nonlinear TS models.

- TAR models allow the model parameters to change according to the value of a threshold variable z_t:

$$y_t = x_t'\beta_j + \varepsilon_t^{(j)}, \quad c_{j-1} < z_t \le c_j$$
$$= \begin{cases} x_t'\beta_1 + \varepsilon_t^{(1)}, & -\infty = c_0 < z_t \le c_1; \\ \vdots & \\ x_t'\beta_k + \varepsilon_t^{(k)}, & c_{k-1} < z_t \le c_k = \infty; \end{cases} \tag{1}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$, and $\varepsilon_t^{(j)}$ are mutually $i.i.d.(0, \sigma_j^2)$.

- The $k-1$ *thresholds* $(c_1, \ldots, c_{k-1})$ divide the domain of the threshold variable z_t into k different regimes. In each different regime, the time series y_t follow a different AR(p) model.

Threshold variable

Two popular types of TAR models are often used in the literature.

- Self-Exciting TAR (SETAR) model: $z_t = y_{t-d}$ – that is, the threshold variable z_t is taken to be a lagged value of the time series itself, where the positive integer d is the delay parameter.

- Momentum TAR (MTAR) model: $z_t = \Delta y_{t-d}$.

Two regime TAR model

Hansen (1997) considers the following two-regime variant of (1)

$$y_t = x_t'\beta_1 I(z_t \leq c) + x_t'\beta_2 I(z_t > c) + \varepsilon_t, \tag{2}$$

where

- $I(A)$ is the indicator function that is equal to 1 if A is true and 0 otherwise;
- $\varepsilon_t \sim iid(0, \sigma^2)$;
- Only one non-trivial threshold c.

Multiple regime with combination variables

- Suppose that the behaviour of the time series y_t not only depends on the value of y_{t-1} relative to some threshold c_1, but also upon the value of y_{t-2} relative to another threshold c_2.

- This gives rise to a four-regime SETAR model

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} \leq c_1 \text{ and } y_{t-2} \leq c_2, \\ \phi_{0,2} + \phi_{1,2} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} \leq c_1 \text{ and } y_{t-2} > c_2, \\ \phi_{0,3} + \phi_{1,3} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} > c_1 \text{ and } y_{t-2} \leq c_2, \\ \phi_{0,4} + \phi_{1,4} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} > c_1 \text{ and } y_{t-2} > c_2. \end{cases} \quad (3)$$

- The model in (3) is referred to as a Nested TAR (NeTAR) model by Astatkie, Watts and Watt (1997).

Extensions of SETAR models

- GJR-GARCH model，Glosten, Jagannathan and Runkle (1993)

$$y_t = u_t, \quad u_t \sim iidN(0, h_t)$$

$$h_t = \omega + \sum_{i=1}^{q}(\alpha_i u_{t-i}^2 + \delta_i d_{t-i} u_{t-i}^2) + \sum_{j=1}^{p} \beta_j h_{t-j}$$

$$d_t = 1, \quad \text{if} \quad u_t \leq 0; \quad \text{else} \quad d_t = 0$$

- Volatility-Switching GARCH (VS-GARCH) model，Fornari and Mele (1996, 1997)

$$\begin{aligned} y_t =& u_t, \quad u_t \sim iidN(0, h_t) \\ h_t =& [\omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}] I(\varepsilon_{t-1} < 0) \\ & + [\xi + \gamma_1 u_{t-1}^2 + \delta_1 h_{t-1}][1 - I(\varepsilon_{t-1} < 0)] \end{aligned}$$

1.2 Smooth Transition Autoregressive (STAR) Models

- Although the TAR models can capture many nonlinear features usually observed in economic and financial time series, sometimes it is counter-intuitive to suggest that the regime switch is abrupt or discontinuous.

- Instead, in some cases it is reasonable to assume that the regime switch happens gradually in a smooth fashion.

- If the discontinuity of the thresholds is replaced by a smooth transition function, TAR models can be generalized to smooth transition autoregressive (STAR) models; see Chan and Tong (1985) and Teräsvirta (1994).

STAR model

If the binary indicator function $I(z_t > c)$ in (2) is replaced by a smooth transition function $0 < G(z_t) < 1$, the model becomes a smooth transition autoregressive (STAR) model

$$y_t = x_t'\beta_1(1 - G(z_t)) + x_t'\beta_2 G(z_t) + \varepsilon_t. \tag{4}$$

where $x_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})'$.

Two transition functions

- Using the logistic function, the transition function can be specified as:

$$G(z_t;\gamma,c)=\frac{1}{1+e^{-\gamma(z_t-c)}},\quad \gamma>0 \tag{5}$$

and the resulting model is called a *Logistic STAR (LSTAR) model*.

  - The *threshold parameter* c can be interpreted as the threshold between the two regimes corresponding to $G(z_t;\gamma,c)=0$ and $G(z_t;\gamma,c)=1$, in the sense that the logistic function changes monotonically from 0 to 1 as y_{t-1} increases, while $G(c;\gamma,c)=0.5$.
  - The *smoothing parameter* γ determines the smoothness of the change in the value of the logistic function, and thus the transition from one regime to the other.

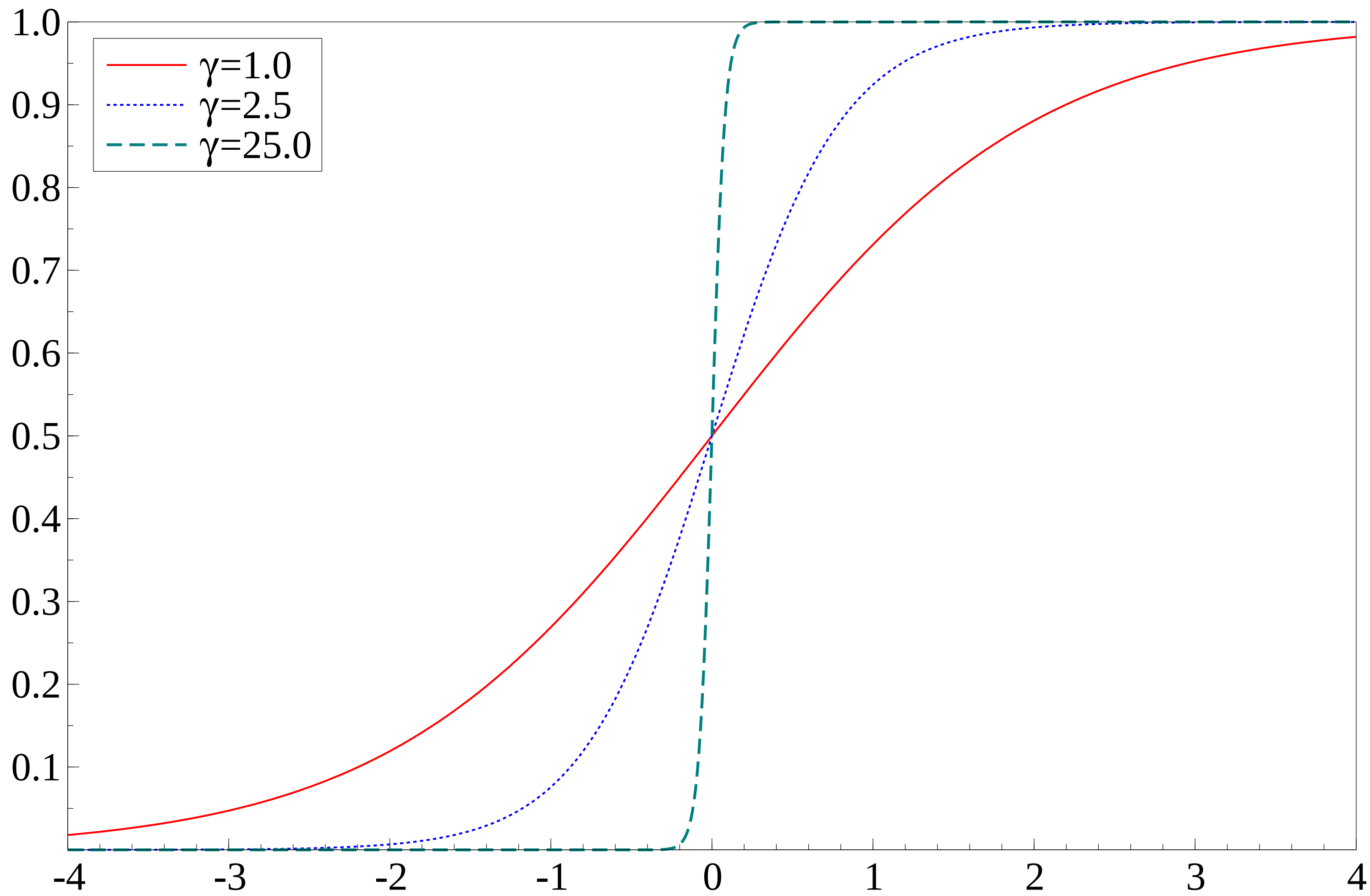

Figure 1 Examples of the logistic function $G(z_t; \gamma, c)$ for various values of the smoothness parameter γ and threshold $c = 0$.

- Using the exponential function, the transition function is specified as:

$$G(z_t;\gamma,c) = 1 - e^{-\gamma(z_t-c)^2}, \quad \gamma > 0, \tag{6}$$

  and the resulting model is recalled as exponential STAR (ESTAR) model.

  – As in LSTAR models, c can be interpreted as the threshold, and γ determines the speed and smoothness of transition.
  – The exponential function is symmetrical and the ESTAR model switches between two regimes smoothly depending on how far the transition variable z_t is from the threshold c.
  – Only the distance between z_t and c matters, but not the sign.

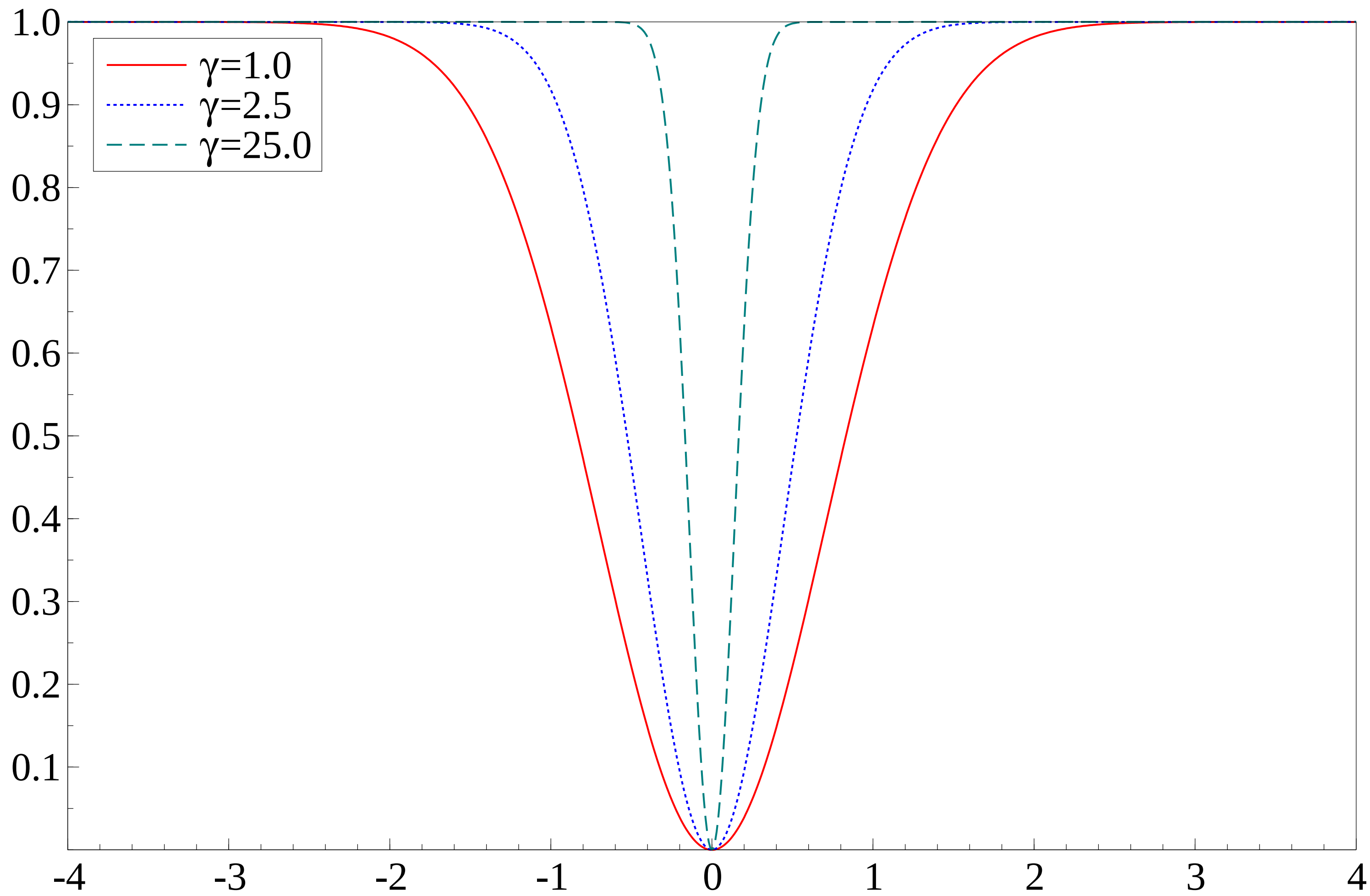

Figure 2 Examples of the exponential function $G(s_t; \gamma, c)$ for various values of the smoothness parameter γ and threshold $c = 0$.

Quadratic logistic function:

A drawback of the exponential function is

- For either $\gamma \to 0$ or $\gamma \to \infty$, the function collapses to a constant, as can be seen in Figure 2.

- Hence, the model becomes linear in both cases and the ESTAR model does not nest a SETAR model as a special case.

This can be remedied by using the quadratic logistic function

$$G(z_t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(z_t-c_1)(z_t-c_2)\}}, \quad c_1 \leq c_2, \gamma > 0, \tag{7}$$

where now $c = (c_1, c_2)'$, as proposed by Jansen and Teräsvirta (1996).

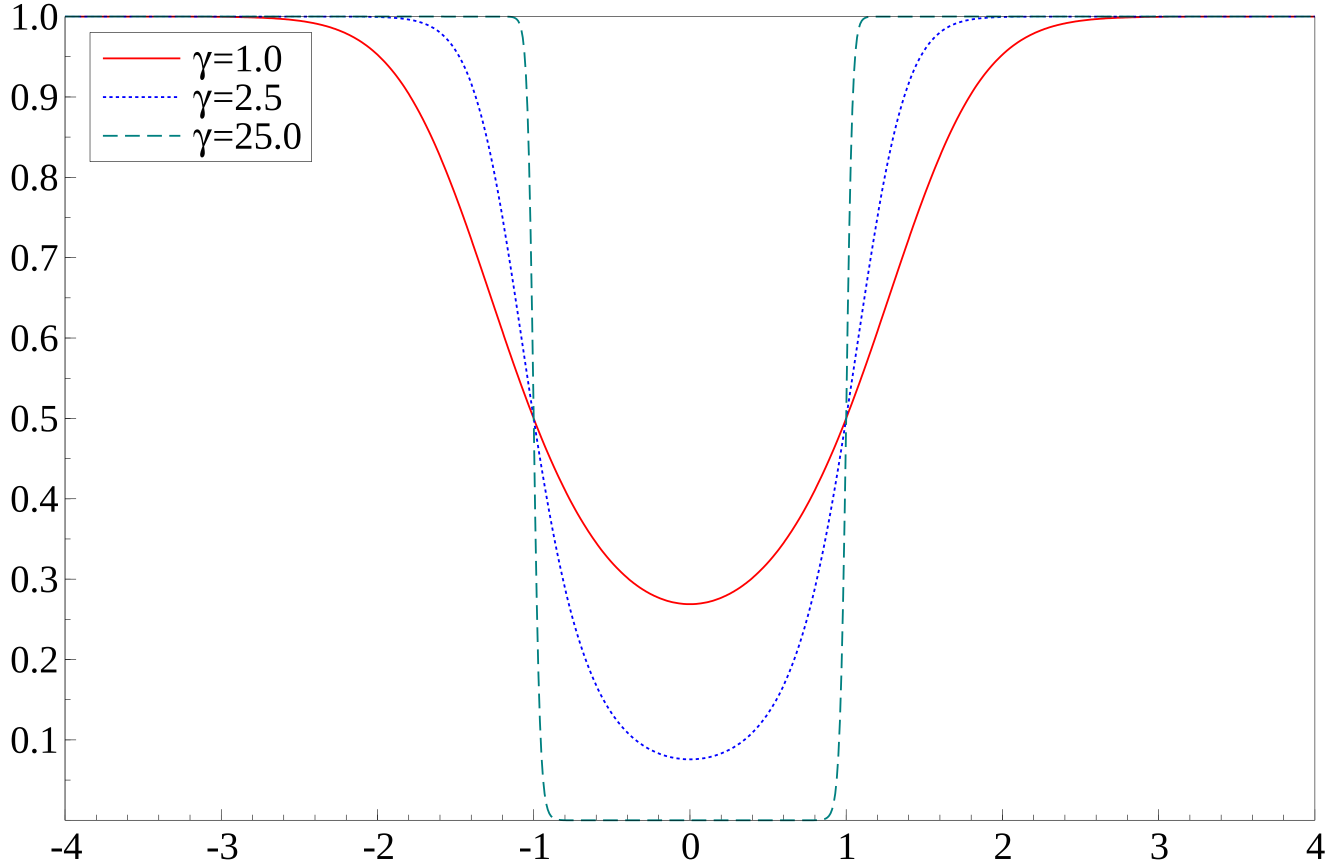

Figure 3 Examples of the quadratic logistic function $G(z_t; \gamma, c)$ for various values of the smoothness parameter γ and $c_1 = -1$ and $c_2 = 1$.

- If $\gamma \to 0$, the model becomes linear, whereas if $\gamma \to \infty$, the function $G(z_t; \gamma, c)$ is equal to 1 for $z_t < c_1$ and $z_t > c_2$ and equal to 0 in between.

- The STAR model with this particular transition function nests a three-regime (SE)TAR model.

- Note that for moderate values of γ, the minimum value taken by the function (7), which is attained for $z_t = (c_1 + c_2)/2$, is not equal to zero; see Figure 3.

Multiple regimes

The STAR models can be extended to multiple regime in a relatively straightforward way. First note that (4) can be rewritten as

$$y_t = x_t'\beta_1 + x_t'(\beta_2 - \beta_1)G(z_t;\gamma,c) + \varepsilon_t, \tag{8}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and $\beta_j = (c, \phi_{j1}, \ldots, \phi_{jp})'$ for $j = 1, 2$. Then, a STAR model with m regimes can be defined as

$$\begin{aligned} y_t = x_t'\beta_1 + x_t'(\beta_2 - \beta_1)G_1(z_t) + x_t'(\beta_3 - \beta_2)G_2(z_t) + \cdots \\ + x_t'(\beta_m - \beta_{m-1})G_{m-1}(z_t) + \varepsilon_t, \end{aligned}$$

where the $G_j(z_t) \equiv G(z_t;\gamma_j,c_j)$, $j = 1, \ldots, m-1$, are logistic functions as in (5) with smoothness parameter γ_j and threshold c_j.

Similar to the four-regime SETAR model given by (3), van Dijk and Franses (1999) propose another multiple regime STAR model. Its representation is

$$
\begin{aligned}
y_t = {} & [x_t'\beta_1(1 - G_1(y_{t-1})) + x_t'\beta_2 G_1(y_{t-1})][1 - G_2(y_{t-2})] \\
& + [x_t'\beta_3(1 - G_1(y_{t-1})) + x_t'\beta_4 G_1(y_{t-1})]G_2(y_{t-2}) + \varepsilon_t,
\end{aligned}
$$

which illustrates the interpretation of nested models perhaps more clearly.

Extensions of STAR models

- Smooth Transition GARCH model

$$y_t = u_t, \quad u_t \sim iidN(0, h_t)$$

$$h_t = \omega + \alpha_1 u_{t-1}^2[1 - G(u_{t-1}; \gamma)] + \delta_1 u_{t-1}^2 G(u_{t-1}; \gamma) + \beta_1 h_{t-1}$$

We can also have another representation

$$y_t = u_t, \quad u_t \sim iidN(0, h_t)$$

$$h_t = \omega + \alpha_1 u_{t-1}^2 + \delta_1 u_{t-1}^2 G(u_{t-1}; \gamma) + \beta_1 h_{t-1}$$

LSTGARCH model: $G(u_{t-1}; \gamma) = \dfrac{1}{1 + \exp(-\gamma u_{t-1})}$.

ESTGARCH model: $G(u_{t-1}; \gamma) = 1 - \exp(-\gamma u_{t-1}^2)$.

- Asymmetric Nonlinear Smooth Transition GARCH (ANSTGARCH) model，Anderson, Nam and Vahid (1999)

$$
\begin{aligned}
y_t &= u_t, \quad u_t \sim iidN(0, h_t) \\
h_t &= (\omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1})[1 - G(u_{t-1}; \gamma)] \\
&\quad + (\xi + \rho_1 u_{t-1}^2 + \delta_1 h_{t-1}) G(u_{t-1}; \gamma)
\end{aligned}
$$

1.3 Markov-Switching Models

Two types of regime switching models

- TAR and STAR models: assume that the regime is determined by the value of z_t relative to a threshold value c.

- Markov-Switching (MS) model: assumes that the regime that occurs at time t cannot be observed, as it is determined by an unobservable process, which we denote as s_t.

Two-regimes MS model

In case of only two regimes, the model with an AR(p) process in both regimes is given by

$$y_t = \begin{cases} x_t'\beta_0 + \varepsilon_t, & \text{if } s_t = 0, \\ x_t'\beta_1 + \varepsilon_t, & \text{if } s_t = 1, \end{cases} \tag{9}$$

or, using an obvious shorthand notation,

$$y_t = x_t'\beta_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2) \tag{10}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$, and $\beta_j = (c_j, \phi_{j1}, \ldots, \phi_{jp})'$.

Note: In this MSAR model, x_t and ε_t are uncorrelated.

The discrete state variable s_t is assumed to be a first-order Markov process with the *transition probabilities*

$$\Pr(s_t = 0|s_{t-1} = 0) = p_{00}, \qquad \Pr(s_t = 0|s_{t-1} = 1) = p_{10},$$
$$\Pr(s_t = 1|s_{t-1} = 0) = p_{01}, \qquad \Pr(s_t = 1|s_{t-1} = 1) = p_{11}.$$

Obviously, for the p_{ij}'s to define proper probabilities, they should be nonnegative, while it should also hold that

$$p_{00} + p_{01} = 1 \quad \text{and} \quad p_{10} + p_{11} = 1.$$

Multiple regime

A MS model with M regimes is obtained by allowing the unobservable Markov chain s_t to take on any one of $M > 2$ different values, each determining a particular regime. That is, the model becomes

$$y_t = x_t' \beta_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2), \tag{11}$$

where s_t is the discrete state variable, taking possible values $\{1, \ldots, M\}$. In general, s_t is assumed to follow a first-order Markov process with the following transition probabilities:

$$p_{ij} \equiv \Pr(s_t = j | s_{t-1} = i), \quad i, j = 1, \ldots, M, \tag{12}$$

which satisfies that $p_{ij} \geq 0$ for $i, j = 1, \ldots, M$ and $\sum_{j=1}^{M} p_{ij} = 1$ for all $i = 1, \ldots, M$.

Endogenous regressors

Consider the following MS model with endogeneous regressors:

$$y_t = x_t'\beta_{s_t} + e_t, \tag{13}$$

$$x_t = Z_t'\delta + \Sigma_v^{1/2} v_t^*, \tag{14}$$

$$\begin{bmatrix} v_t^* \\ e_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_K & \rho_{s_t}\sigma_{e,s_t} \\ \rho_{s_t}'\sigma_{e,s_t} & \sigma_{e,s_t}^2 \end{bmatrix} \right), \tag{15}$$

- y_t is a dependent variable;
- x_t is a $K \times 1$ vector of endogenous variables;
- $Z_t = I_K \otimes z_t$, with z_t being an $L \times 1$ $(L \geq K)$ vector of instrumental variables;
- ρ_{s_t} is an $K \times 1$ vector of correlation coefficients.

As shown in Kim (2004), the Cholesky decomposition of the covariance matrix of $[v_t^* \quad e_t]'$ in Eq. (15) results in the following representation:

$$\begin{bmatrix} v_t^* \\ e_t \end{bmatrix} = \begin{bmatrix} I_K & 0 \\ \rho'_{s_t}\sigma_{e,s_t} & \sqrt{1-\rho'_{s_t}\rho_{s_t}}\sigma_{e,s_t} \end{bmatrix} \begin{bmatrix} v_t^* \\ \eta_t \end{bmatrix},$$

It is easy to find that $e_t = \rho'_{s_t}\sigma_{e,s_t}v_t^* + \sqrt{1-\rho'_{s_t}\rho_{s_t}}\sigma_{e,s_t}\eta_t$.

A key to successful estimation of the model is in the orthogonal decomposition of e_t, which allows one to rewrite Equation (13) as

$$y_t = x_t'\beta_{s_t} + \gamma'_{s_t}v_t^* + w_t, \quad w_t \mid s_t \sim i.i.d.N(0, \sigma^2_{w,s_t}),$$

where $\gamma_{s_t} = \rho_{s_t}\sigma_{e,s_t}$ and $\sigma^2_{w,s_t} = (1-\rho'_{s_t}\rho_{s_t})\sigma^2_{e,s_t}$. Conditional on x_t and s_t, the distribution term w_t is independent of either x_t and v_t^*.

See for example

(1) Kim, C.-J., 2004, “Markov-Switching Models with Endogenous Explanatory Variables”, Journal of Econometrics, 122, 127-136.

(2) Kim, C.-J., 2006, “Time-Varying Parameter Models with Endogenous Regressors”, Economics Letters, 91, 21-26.

(3) Kim, C.-J., 2009, “Markov-Switching Models with Endogenous Explanatory Variables II: A Two-Step MLE Procedure”, Journal of Econometrics, 148, 46-55.

Extensions of MS models

- MSM model, Hamilton (1989)'s method for measuring business cycle

$$y_t = \mu_{s_t} + \phi_1(y_{t-1} - \mu_{s_{t-1}}) + \cdots + \phi_p(y_{t-p} - \mu_{s_{t-p}}) + \varepsilon_t$$

- Time-varying transition probability (TVTP) model, Filardo (1994)、Filardo (1998)

$$\begin{aligned} p_{ij,t} &= \Pr(s_t = j | s_{t-1} = i, z_{t-1}) \\ &= \frac{\exp(\lambda_{ij,0} + z'_{t-1}\lambda_{ij,1})}{1 + \exp(\lambda_{ij,0} + z'_{t-1}\lambda_{ij,1})} \end{aligned}$$

- Switching ARCH model (SW-ARCH), Hamilton and Susmel (1994), Markov jump and volatility cluster behaviors

$$y_t = u_t, \quad u_t \sim i.i.d.(0, h_t)$$

$$\frac{h_t}{\gamma_{s_t}} = \alpha_0 + \frac{\alpha_1}{\gamma_{s_{t-1}}} u_{t-1}^2 + \ldots + \frac{\alpha_r}{\gamma_{s_{t-q}}} u_{t-q}^2$$

Practically, one of γ should be set as 1, i.e., $\gamma_1 = 1$.

Strong convergence using analytical derivatives
Log-likelihood value = -11349.2

	Coef.Est	Std.Err	t-value	t-prob
p11	0.996998	0.001896	525.827360	0.000000
p22	0.994224	0.002813	353.441956	0.000000
phi0	0.039767	0.013937	2.853406	0.004339
alpha0	0.839673	0.066695	12.589662	0.000000
alpha1	0.020462	0.014448	1.416294	0.156737
alpha2	0.116608	0.022096	5.277375	0.000000
alpha3	0.104663	0.022636	4.623696	0.000004
alpha4	0.068243	0.020768	3.285976	0.001022
alpha5	0.052686	0.020541	2.564924	0.010342
gamma2	4.282024	0.378576	11.310860	0.000000
lever	0.065051	0.029827	2.180921	0.029225
t-nu	4.420363	0.280562	15.755369	0.000000

Table 1 Parameter estimates of SWARCH model for SH from 1995 to 2022.

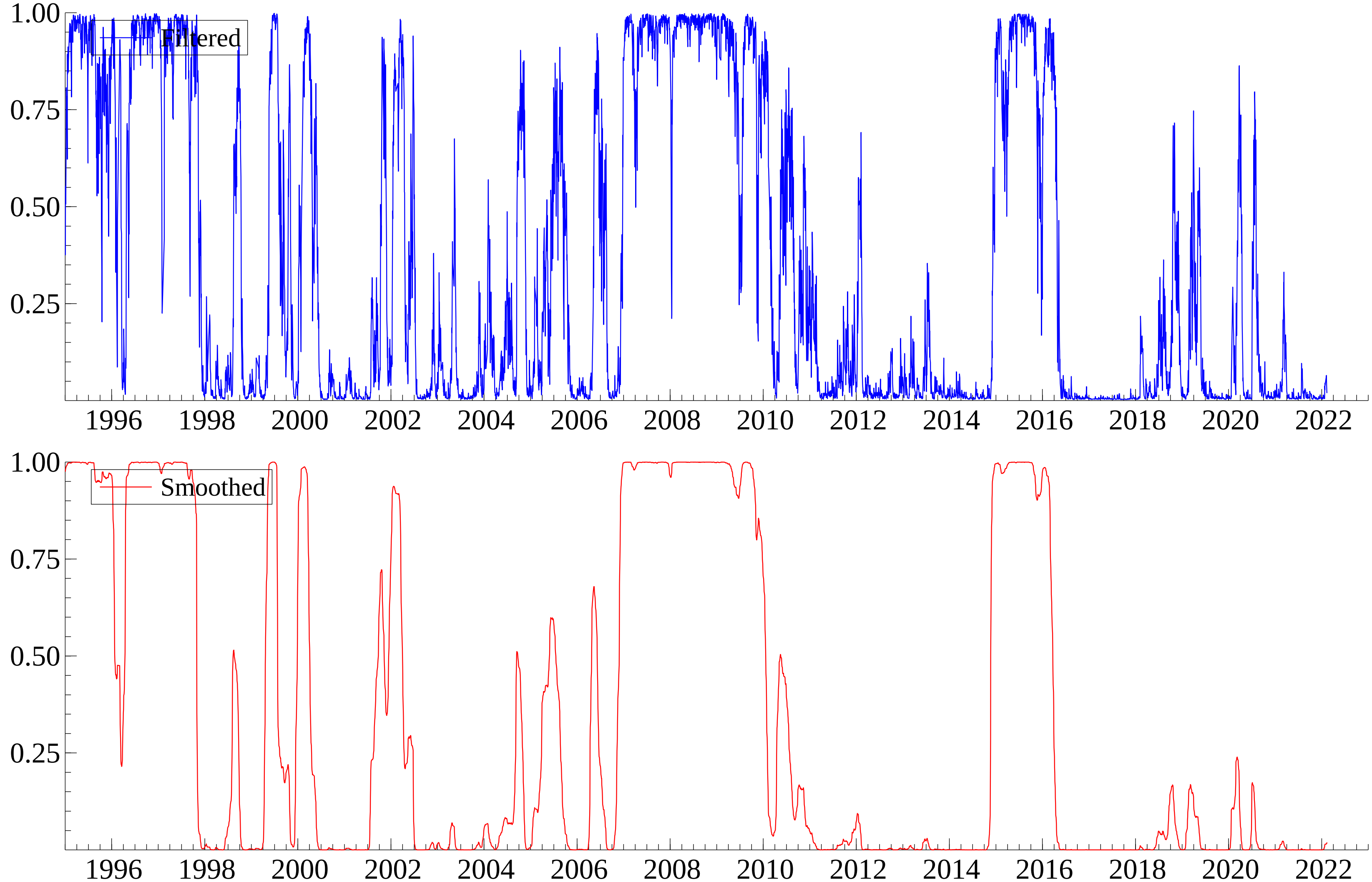

Figure 4 Regime probability of SWARCH model for SH from 1995 to 2022.

- Bivariate SWARCH model or BISWARCH, Ramchand and Susmel (1998) contagion effects for financial markets

- Markov switching GARCH (MS-GARCH) model, Gary (1996), Klaasen (2002)

- Markov switching stochastic volatility (MSSV) model, So, Lam, and Li (1998), Kalimipalli and Susmel (2004), Lichen Sun (2005), Hwang, Satchell, and Valls (2007)

$$
\begin{aligned}
y_t &= \exp(h_t/2)\varepsilon_t \\
h_t &= \mu_{s_t} + \phi h_{t-1} + \eta_t \\
\varepsilon_t &\sim iidN(0,1), \quad \eta_t \sim iidN(0,\sigma_\eta^2)
\end{aligned}
$$

1.4 Model specifications

An empirical specification procedure for SETAR, STAR and MS models that follows this approach consists of the following steps:

(1) specify an appropriate linear AR model of order p [AR(p)] for the time series under investigation;

(2) test the null hypothesis of linearity against the alternative of SETAR-, STAR- and/or MS-type nonlinearity; for the SETAR and STAR models, this step also consists of selecting the appropriate variable that determines the regimes;

(3) estimate the parameters in the selected model;

(4) evaluate the model using diagnostic tests;

(5) modify the model if necessary;

(6) use the model for descriptive or forecasting purposes.

2 Estimation of Nonlinear Models

2.1 SETAR models

The parameters of interest in the 2-regime SETAR model can conveniently be estimated by sequential conditional least squares. [1]

Again the 2-regime SETAR model

$$y_t = x_t'\phi_1(1 - I(z_t > c)) + x_t'\phi_2 I(z_t > c) + \varepsilon_t.$$

[1] Under the additional assumption that the ε_ts are normally distributed, the resulting estimates are equivalent to maximum likelihood estimates.

- When the threshold c is fixed, the model is linear in the remaining parameters. Estimates of $\phi = (\phi_1', \phi_2')'$ are then easily obtained by OLS as

$$\hat{\phi}(c) = \left(\sum_{t=1}^{n} x_t(c) x_t(c)' \right)^{-1} \left(\sum_{t=1}^{n} x_t(c) y_t \right), \tag{16}$$

where $x_t(c) = (x_t' I[y_{t-1} \leq c], x_t' I[y_{t-1} > c])'$ and the notation $\hat{\phi}(c)$ is used to indicate that the estimate of ϕ is conditional upon c.

- The corresponding residuals are denoted as

$$\hat{\varepsilon}_t(c) = y_t - \hat{\phi}(c)' x_t(c)$$

with variance $\hat{\sigma}^2(c) = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t(c)^2$.

- The LS estimate of c can be obtained by minimizing this residual variance, that is

$$\hat{c} = \operatorname*{argmin}_{c \in C} \hat{\sigma}^2(c), \tag{17}$$

where C denotes the set of all allowable threshold values. The minimization problem (17) can be solved by means of direct search.

- The final estimates of the autoregressive parameters are given by

$$\hat{\phi} = \hat{\phi}(\hat{c}),$$

while the residual variance is estimated as

$$\hat{\sigma}^2 = \hat{\sigma}^2(\hat{c}).$$

(1) Choice of the set C

A popular choice for C is to require that each regime contains at least a (pre-specified) fraction π_0 of the observations, that is,

$$C = \{c | y_{([\pi_0(n-1)])} \leq c \leq y_{([(1-\pi_0)(n-1)])}\}, \tag{18}$$

where $y_{(0)}, \ldots, y_{(n-1)}$ denote the order statistics of the threshold variable y_{t-1}, $y_{(0)} \leq \cdots y_{(n-1)}$, and $[\cdot]$ denotes integer part. A safe choice for π_0 appears to be 0.15.

(2) Choosing the threshold variable

Generally in **SETAR models**, lagged endogenous variables y_{t-d} is used as candidate threshold variables.

The delay parameter d can be estimated along with the other parameters in the model, by performing the above calculations for various choices of d (say, $d \in \{1, \ldots, d^*\}$ for some upper bound d^*), and estimate d as the value that minimizes the residual variance.

So the *grid search* in (17) is augmented with a search over d – that is, the minimization problem becomes

$$(\hat{c}, \hat{d}) = \underset{c \in C, d \in D}{\operatorname{argmin}} \; \hat{\sigma}^2(c, d), \tag{19}$$

where $D = \{1, \ldots, d^*\}$ and $\hat{\sigma}^2(c, d)$ is the estimate of residual variance.

If one wants to allow for an exogenous threshold variable q_t, a similar procedure can be followed.

2.2 STAR models

Estimation method: nonlinear least squares (NLS)

The parameters $\theta = (\phi_1', \phi_2', \gamma, c)'$ can be estimated as

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q_n(\theta) = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} [y_t - F(x_t; \theta)]^2, \tag{20}$$

where $F(x_t; \theta)$ is the skeleton of the model, that is,

$$F(x_t; \theta) \equiv \phi_1' x_t (1 - G(y_{t-1}; \gamma, c)) + \phi_2' x_t G(y_{t-1}; \gamma, c).$$

Under the assumption that the errors ε_t are normally distributed, NLS is equivalent to maximum likelihood. Otherwise, the NLS estimates can be interpreted as quasi-maximum likelihood estimates.

Under certain regularity conditions, the NLS estimates are consistent and asymptotically normal, that is,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to N(0, C), \tag{21}$$

where θ_0 denotes the true parameter values.

(1) Starting values

- Given the parameters γ and c in the transition function, the STAR model is linear in AR parameters ϕ_1 and ϕ_2, similar to SETAR model.

- Conditional upon γ and c, the OLS estimates of $\phi = (\phi_1', \phi_2')'$

$$\hat{\phi}(\gamma, c) = \left(\sum_{t=1}^{n} x_t(\gamma, c) x_t(\gamma, c)' \right)^{-1} \left(\sum_{t=1}^{n} x_t(\gamma, c) y_t \right), \qquad (22)$$

where $x_t(\gamma, c) = (x_t'(1 - G(y_{t-1}; \gamma, c)), x_t' G(y_{t-1}; \gamma, c))'$.

- The residuals

$$\hat{\varepsilon}_t = y_t - \hat{\phi}(\gamma, c)' x_t(\gamma, c)$$

with associated variance $\hat{\sigma}^2(\gamma, c) = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2(\gamma, c)$.

Grid search

A convenient method to obtain sensible starting values for the nonlinear optimization algorithm is to perform a two-dimensional grid search over γ and c and select those parameter estimates which render the smallest estimate for the residual variance $\hat{\sigma}^2(\gamma, c)$.

(2) Concentrating the sum of squares function

As suggested by Leybourne, Newbold and Vougas (1998), another way to simplify the estimation problem is to concentrate the sum of squares function.

Owing to the fact that the STAR model is linear in the autoregressive parameters for given values of γ and c, the sum of squares function $Q_n(\theta)$ can be concentrated with respect to ϕ_1 and ϕ_2 as

$$Q_n(\gamma, c) = \sum_{t=1}^{n} (y_t - \phi(\gamma, c)' x_t(\gamma, c))^2 \longrightarrow \min . \tag{23}$$

This reduces the dimensionality of the NLS estimation problem considerably, as the sum of squares function as given in (23) needs to be minimized with respect to the two parameters γ and c only.

(3) The estimate of γ

- The estimate of γ is rather imprecise in general and often appears to be insignificant when judged by its t-statistic.

- One reason is that for large values of γ, the shape of the logistic function (5) changes only little. Hence, to obtain an accurate estimate of γ one needs many observations in the immediate neighbourhood of the threshold c.

- The insignificance of the estimate of γ should not be interpreted as evidence against the presence of STAR-type nonlinearity.

- This should be assessed by means of different diagnostics, some of which are discussed below.

2.3 Markov switching models

The parameters in the MS model can be estimated using maximum likelihood techniques. However, due to the fact that the state variable s_t is not observed, the estimation problem is highly nonstandard.

Consider the estimation of the MS model (11). Again,

$$y_t = x_t' \beta_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2).$$

The parameters to be estimated include:
(1) AR coefficients in the different regimes,
(2) transition probability parameters,
(3) the probabilities with which each state occurs at each point in time.

First, the conditional likelihood function at date t is

$$f(y_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{M} f(y_t|s_t = j, \mathcal{F}_{t-1}) \Pr(s_t = j|\mathcal{F}_{t-1}) \tag{24}$$

following the total probability. It is obviously that the density of $f(y_t|s_t = j, \mathcal{F}_{t-1})$ is known by

$$f(y_t|s_t = j, \mathcal{F}_{t-1}) = \left(2\pi\sigma_j^2\right)^{-1/2} \exp\left\{-\frac{(y_t - x_t'\beta_j)^2}{2\sigma_j^2}\right\} \tag{25}$$

In the equation (24), the one-step-ahead-forecast density

$$\Pr(s_t = j|\mathcal{F}_{t-1}) = \sum_{i=1}^{M} p_{ij} \Pr(s_{t-1} = i|\mathcal{F}_{t-1}) \tag{26}$$

where the transition probability $p_{ij} = \Pr(s_t = j|s_{t-1} = i)$.

The probability $\Pr(s_t = j|\mathcal{F}_t)$ is called "filtered probability". Using the Bayes rule, it can be computed as

$$\begin{aligned}
\Pr(s_t = j|\mathcal{F}_t) &= \Pr(s_t = j|y_t, \mathcal{F}_{t-1}) && (27)\\
&= \frac{f(s_t = j, y_t|\mathcal{F}_{t-1})}{f(y_t|\mathcal{F}_{t-1})} \\
&= \frac{f(y_t|s_t = j, \mathcal{F}_{t-1}) \Pr(s_t = j|\mathcal{F}_{t-1})}{f(y_t|\mathcal{F}_{t-1})} && (28)
\end{aligned}$$

where $f(y_t|s_t = j, \mathcal{F}_{t-1})$ is given by (25), $f(y_t|\mathcal{F}_{t-1})$ is given by (24), and $\Pr(s_t = j|\mathcal{F}_{t-1})$ is given by (26).

So given the input probability $\Pr(s_{t-1} = i|\mathcal{F}_{t-1})$ at time $t-1$, $i = 1, \ldots, M$, we can implement the output probability $\Pr(s_t = j|\mathcal{F}_t)$ at time t.

(1) Filtering Algorithm: called the Hamilton filter

Algorithm 1. *For $t = 1, 2, \ldots, T$, the following algorithm is carried out recursively.*

1. *For $j = 1, \ldots, M$, the one-step-forecast-density of discrete variable s_t is*

$$\Pr(s_t = j|\mathcal{F}_{t-1}) = \sum_{i=1}^{M} p_{ij} \Pr(s_{t-1} = i|\mathcal{F}_{t-1})$$

or in vector-form

$$\xi_{t|t-1} = P\xi_{t-1|t-1}$$

2. *For $j = 1, \ldots, M$, the filtered probabilities are given by*

$$\Pr(s_t = j|\mathcal{F}_t) = \frac{f(y_t|s_t = j, \mathcal{F}_{t-1}) \Pr(s_t = j|\mathcal{F}_{t-1})}{\sum_{k=1}^{M} f(y_t|s_t = k, \mathcal{F}_{t-1}) \Pr(s_t = k|\mathcal{F}_{t-1})}$$

or in vector-form

$$\xi_{t|t} = \frac{\boldsymbol{f}_t \odot \xi_{t|t-1}}{\mathbf{1}'(\boldsymbol{f}_t \odot \xi_{t|t-1})}$$

In this algorithm, we can get the conditional likelihood function as follow

$$f(y_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{M} f(y_t|s_t = j, \mathcal{F}_{t-1}) \Pr(s_t = j|\mathcal{F}_{t-1})$$

or in vector-form

$$f(y_t|\mathcal{F}_{t-1}) = \mathbf{1}'(\boldsymbol{f}_t \odot \xi_{t|t-1})$$

Using this conditional likelihood function, the MLE can be easily constructed.

(2) Smoothing Algorithm

To conduct the smoothed probabilities, we first consider the joint probability between $s_t = j$ and $s_{t+1} = k$, that is

$$\begin{aligned}
&\Pr(s_t = j, s_{t+1} = k|\mathcal{F}_T) \\
=&\Pr(s_{t+1} = k|\mathcal{F}_T)\Pr(s_t = j|s_{t+1} = k, \mathcal{F}_T) \\
=&\Pr(s_{t+1} = k|\mathcal{F}_T)\Pr(s_t = j|s_{t+1} = k, \mathcal{F}_t) \\
=&\Pr(s_{t+1} = k|\mathcal{F}_T)\frac{\Pr(s_t = j, s_{t+1} = k|\mathcal{F}_t)}{\Pr(s_{t+1} = k|\mathcal{F}_t)} \\
=&\Pr(s_{t+1} = k|\mathcal{F}_T)\frac{p_{jk}\Pr(s_t = j|\mathcal{F}_t)}{\Pr(s_{t+1} = k|\mathcal{F}_t)} \qquad (29)
\end{aligned}$$

From the second line to the third line is followed by Kim (1994).

Sum all over k for the joint probabilities, we get

$$\Pr(s_t = j|\mathcal{F}_T) = \sum_{k=1}^{M} \Pr(s_t = j, s_{t+1} = k|\mathcal{F}_T)$$

or in vector-form

$$\xi_{t|n} = \xi_{t|t} \odot (P'[\xi_{t+1|n} \div \xi_{t+1|t}])$$

This is called the "smoothed probabilities".

(3) Steady-State Probabilities (Ergodic Probability)

Let the transition probabilities for a first-order, M-state Markov-switching process s_t be represented as

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{M1} \\ p_{12} & p_{22} & \cdots & p_{M2} \\ \vdots & \vdots & \ddots \vdots & \\ p_{1M} & p_{2M} & \cdots & p_{MM} \end{pmatrix}, \quad \sum_{j=1}^{M} p_{ij} = 1, \forall i. \tag{30}$$

Hence, we have

$$i'_M P = i'_M$$

with $i_M = \mathbf{1} = (1, 1, \ldots, 1)'$.

If let π be a vector of $M \times 1$ steady-state probabilities, we have

$$\pi_t = \begin{pmatrix} \Pr(s_t = 1) \\ \Pr(s_t = 2) \\ \vdots \\ \Pr(s_t = M) \end{pmatrix} = \begin{pmatrix} \pi_{1t} \\ \pi_{2t} \\ \vdots \\ \pi_{Mt} \end{pmatrix}, \quad i_M' \pi_t = 1.$$

By definition of steady-state probabilities, $\pi_{t+1} = P\pi_t$ and $\pi_{t+1} = \pi_t$, thus

$$\pi_t = P\pi_t \quad \Rightarrow \quad (I_M - P)\pi_t = \mathbf{0}_{M\times 1}.$$

Combining $i_M' \pi_t = 1$,

$$\begin{pmatrix} I_M - P \\ i_M' \end{pmatrix} \pi_t = \begin{pmatrix} \mathbf{0}_{M\times 1} \\ 1 \end{pmatrix}, \quad \text{or} \quad A\pi_t = \begin{pmatrix} \mathbf{0}_{M\times 1} \\ 1 \end{pmatrix}$$

Thus,

$$\pi_t = (A'A)^{-1}A' \begin{pmatrix} \mathbf{0}_{M\times 1} \\ 1 \end{pmatrix}.$$

That is, π_t is the last column of the matrix $(A'A)^{-1}A'$.

(4) Expected Duration

Denote D as the duration of state j,

$$\begin{aligned}
D &= 1, \quad \text{if } s_t = j\text{, and } s_{t+1} \neq j\text{; then } \Pr(D=1) = 1 - p_{jj} \\
D &= 2, \quad \text{if } s_t = s_{t+1} = j\text{, and } s_{t+2} \neq j\text{; then } \Pr(D=2) = p_{jj}(1 - p_{jj}) \\
D &= 3, \quad \text{if } s_t = s_{t+1} = s_{t+2} = j\text{, and } s_{t+3} \neq j\text{; then } \Pr(D=3) = p_{jj}^2(1 - p_{jj}) \\
\vdots &= \vdots
\end{aligned}$$

Then, the expected duration of regime j can be derived as

$$
\begin{aligned}
E(D) &= \sum_{j=1}^{\infty} j \Pr(D = j) \\
&= 1 \times (1 - p_{jj}) + 2 \times p_{jj}(1 - p_{jj}) + 3 \times p_{jj}^2(1 - p_{jj}) + \cdots \\
&= (1 - p_{jj})(1 + 2p_{jj} + 3p_{jj}^2 + \cdots) \\
&= (1 - p_{jj}) \frac{d}{dp_{jj}} \left(\frac{p_{jj}}{1 - p_{jj}} \right) \\
&= \frac{1}{1 - p_{jj}}
\end{aligned}
$$

(5) Expectation Maximization (EM) algorithm

Following Hamilton (1990), the MLE of transition probabilities are

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} \Pr(s_t = j, s_{t-1} = i | \mathcal{F}_T; \hat{\theta})}{\sum_{t=2}^{T} \Pr(s_{t-1} = i | \mathcal{F}_T; \hat{\theta})}, \tag{31}$$

where $\hat{\theta}$ denotes the MLE of θ.

It is also shown that these satisfy the first-order conditions (FOC)

$$\sum_{t=1}^{T} (y_t - x_t' \hat{\beta}_j) x_t \Pr(s_t = j | \mathcal{F}_T; \hat{\theta}) = 0, \quad j = 1, 2, \tag{32}$$

and

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{2} (y_t - x_t' \hat{\beta}_j)^2 \Pr(s_t = j | \mathcal{F}_T; \hat{\theta}), \tag{33}$$

Notice that (32) implies that $\hat{\beta}_j$ is the estimate corresponding to a weighted least squares regression of y_t on x_t, with weights given by the square root of the smoothed probability of regime j occurring. Hence, the estimates $\hat{\beta}_j$ can be obtained as

$$\hat{\beta}_j = \left(\sum_{t=1}^{T} x_t(j) x_t(j)' \right)^{-1} \left(\sum_{t=1}^{T} x_t(j) y_t(j) \right), \tag{34}$$

where

$$y_t(j) = y_t \sqrt{\Pr(s_t = j | \mathcal{F}_T; \hat{\theta})},$$
$$x_t(j) = x_t \sqrt{\Pr(s_t = j | \mathcal{F}_T; \hat{\theta})}.$$

Finally, the ML estimate of the residual variance is obtained using (33) as the mean of the squared residuals from the two WLS regressions.

Algorithm 2 (Expectation Maximization)**.** *1. Given starting values for the parameter vector* $\hat{\theta}^{(0)}$*, first compute the smoothed regime probabilities.*

2. The smoothed regime probabilities $\hat{\xi}_{t|n}$ *are combined with the initial estimates of the transition probabilities* $\hat{p}_{ij}^{(0)}$ *to obtain new estimates of the transition probabilities* $\hat{p}_{ij}^{(1)}$ *from* (31)*.*

3. (34) *and* (33) *can be used to obtain a new set of estimates of the autoregressive parameters and the residual variance. Combined with the new estimates of the transition probabilities, this gives a new set of estimates for all parameters in the model,* $\hat{\theta}^{(1)}$*.*

4. Iterating this procedure renders estimates $\hat{\theta}^{(2)}, \hat{\theta}^{(3)}, \ldots$ *and this can be continued until convergence occurs.*

(6) Transition probability in MSM model

Consider a two-regime MSM model proposed by Hamilton (1989),

$$y_t = \mu_{s_t} + \phi_1(y_{t-1} - \mu_{s_{t-1}}) + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, \sigma^2).$$

In this model, s_t takes values 0 or 1 and y_t depends on both s_t and s_{t-1}. Moreover, the transition probability matrix for s_t is defined as

$$P = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}.$$

However, it is uneasy to express the likelihood function directly as done previously.

We now introduce another way by defining a new state variable, i.e.,

$$s_t^* = \begin{cases} 1, & \text{if } s_t = 0 \text{ and } s_{t-1} = 0; \\ 2, & \text{if } s_t = 1 \text{ and } s_{t-1} = 0; \\ 3, & \text{if } s_t = 0 \text{ and } s_{t-1} = 1; \\ 4, & \text{if } s_t = 1 \text{ and } s_{t-1} = 1. \end{cases} \tag{35}$$

Then the MSM model can be represented by

$$y_t = \mu_{s_t} + \phi_1 \mu_{s_{t-1}} + \phi_1 y_{t-1} + \varepsilon_t \equiv c_{s_t^*} + \phi_1 y_{t-1} + \varepsilon_t$$

with a new transition probability matrix

$$P^* = [\Pr(s_t^* = j | s_{t-1}^* = i)],$$

where the transition probability can be represented by

$$
\begin{array}{l}
\hline
\Pr(s_t^* = 1 | s_{t-1}^* = 1) = \Pr(s_t = 0, s_{t-1} = 0 | s_{t-1} = 0, s_{t-2} = 0) = p_{00} \\
\Pr(s_t^* = 2 | s_{t-1}^* = 1) = \Pr(s_t = 1, s_{t-1} = 0 | s_{t-1} = 0, s_{t-2} = 0) = 1 - p_{00} \\
\Pr(s_t^* = 3 | s_{t-1}^* = 1) = \Pr(s_t = 0, s_{t-1} = 1 | s_{t-1} = 0, s_{t-2} = 0) = 0 \\
\Pr(s_t^* = 4 | s_{t-1}^* = 1) = \Pr(s_t = 1, s_{t-1} = 1 | s_{t-1} = 0, s_{t-2} = 0) = 0 \\
\hline
\Pr(s_t^* = 1 | s_{t-1}^* = 2) = \Pr(s_t = 0, s_{t-1} = 0 | s_{t-1} = 1, s_{t-2} = 0) = 0 \\
\Pr(s_t^* = 2 | s_{t-1}^* = 2) = \Pr(s_t = 1, s_{t-1} = 0 | s_{t-1} = 1, s_{t-2} = 0) = 0 \\
\Pr(s_t^* = 3 | s_{t-1}^* = 2) = \Pr(s_t = 0, s_{t-1} = 1 | s_{t-1} = 1, s_{t-2} = 0) = 1 - p_{11} \\
\Pr(s_t^* = 4 | s_{t-1}^* = 2) = \Pr(s_t = 1, s_{t-1} = 1 | s_{t-1} = 1, s_{t-2} = 0) = p_{11} \\
\hline
\end{array}
$$

$$\Pr(s_t^* = 1|s_{t-1}^* = 3) = \Pr(s_t = 0, s_{t-1} = 0|s_{t-1} = 0, s_{t-2} = 1) = p_{00}$$

$$\Pr(s_t^* = 2|s_{t-1}^* = 3) = \Pr(s_t = 1, s_{t-1} = 0|s_{t-1} = 0, s_{t-2} = 1) = 1 - p_{00}$$

$$\Pr(s_t^* = 3|s_{t-1}^* = 3) = \Pr(s_t = 0, s_{t-1} = 1|s_{t-1} = 0, s_{t-2} = 1) = 0$$

$$\Pr(s_t^* = 4|s_{t-1}^* = 3) = \Pr(s_t = 1, s_{t-1} = 1|s_{t-1} = 0, s_{t-2} = 1) = 0$$

$$\Pr(s_t^* = 1|s_{t-1}^* = 4) = \Pr(s_t = 0, s_{t-1} = 0|s_{t-1} = 1, s_{t-2} = 1) = 0$$

$$\Pr(s_t^* = 2|s_{t-1}^* = 4) = \Pr(s_t = 1, s_{t-1} = 0|s_{t-1} = 1, s_{t-2} = 1) = 0$$

$$\Pr(s_t^* = 3|s_{t-1}^* = 4) = \Pr(s_t = 0, s_{t-1} = 1|s_{t-1} = 1, s_{t-2} = 1) = 1 - p_{11}$$

$$\Pr(s_t^* = 4|s_{t-1}^* = 4) = \Pr(s_t = 1, s_{t-1} = 1|s_{t-1} = 1, s_{t-2} = 1) = p_{11}$$

(7) Determination of the number of regimes

- In Markov-switching models, Akaike's information criterion (AIC) may mislead the users into selecting too many states.

- Smith et al. (2006) suggest using the Markov switching criterion (MSC) to select the number of states and variables simultaneously.

Using the Kullback-Leibler (KL) divergence, Smith et al. (2006) derived the Markov switching criterion

$$MSC = -2\log L + \sum_{i=1}^{M} \frac{T_i(T_i + \lambda_i K)}{\delta_i T_i - \lambda_i K - 2} \tag{36}$$

where M is the number of states and K is the number of the regressors in the model, T_i is the sum of smoothed probabilities of being in the regime i and so $\sum_{i=1}^{M} T_i = T$, where T is the number of the effective observations used in estimation.

The second term in (36) is a penalty for the complexity of a model. Smith et al. (2006) recommended setting $\delta_i = 1$ and $\lambda_i = N$ (also $\lambda_i = 1$ or $\lambda_i = N^2$ is another possible choice but they showed that $MSC_{\lambda=1}$ and $MSC_{\lambda=N^2}$ perform worse than when both the sample size is small and the signal is weak).

See detailedly in Smith, A., Naik, P., Tsai, C., 2006, "Markov-Switching Model Selection Using Kullback-Leibler Divergence", Journal of Econometrics, 134, 553-577.

3 Nonlinearity Tests

3.1 Threshold test

Consider testing the Null Hypothesis that a time series y_t being a linear AR model $y_t = \phi_0 + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t$ against the Alternative Hypothesis that a 2-regime self-exciting TAR (SETAR) model

$$y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + \cdots + \phi_p^{(1)} y_{t-p} + \varepsilon_t^{(1)}, & \text{if } y_{t-d} < c_1; \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + \cdots + \phi_p^{(2)} y_{t-p} + \varepsilon_t^{(2)}, & \text{if } y_{t-d} \geq c_1. \end{cases}$$

where c_1 is the threshold.

For a given realization $\{y_t\}_{t=1}^{T}$ and assuming normality, the likelihood functions for the test are defined by

- ℓ_0 for linear AR model is the log-likelihood function evaluated at maximum-likelihood estimates of $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_p)'$ and σ^2;

- $\ell_1(c_1)$ for TAR model is the log-likelihood function evaluated at maximum-likelihood estimates of $\boldsymbol{\phi}_i = (\phi_0^{(i)}, \ldots, \phi_p^{(i)})'$ and σ_i^2 for $i = 1, 2$ conditioned on knowing the thresholds c_1.

Then the log-likelihood ratio LR is defined as

$$LR = \ell_1(c_1) - \ell_0.$$

The difficulty is that the threshold c_1 under the null hypothesis of linear model is not defined (so it is often called as a *nuisance parameter*). This results in that the asymptotic distribution of the likelihood ratio is very different from that of the conventional likelihood ratio statistics.

Hansen's sup-LR test

Hansen (1997) considers the following 2-regime variant

$$y_t = x_t'\beta_1 I(y_{t-d} \leq c) + x_t'\beta_2 I(y_{t-d} > c) + \varepsilon_t, \tag{37}$$

where $I(A)$ is the indicator function that is equal to 1 if A is true and 0 otherwise, $\varepsilon_t \sim iid(0, \sigma^2)$, and there is only one non-trivial threshold c.

To test the null hypothesis of linear model against the alternative hypothesis of 2-regime SETAR model, the likelihood ratio test assuming normally distributed errors can be used:

$$F(c) = \frac{RSS_0 - RSS_1}{\hat{\sigma}_1^2(c)} = n\frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2(c)}{\hat{\sigma}_1^2(c)}, \tag{38}$$

where RSS_0 is from linear model, RSS_1 is from 2-regime SETAR model given the threshould c, n is the effective sample size, $\hat{\sigma}_0^2$ is residual variance of linear model, and $\hat{\sigma}_1^2(c)$ is residual variance of nonlinear model.

Since the threshold c is usually unknown, Hansen (1997) suggests to compute the following $\sup\!-\!LR$ test:

$$F_s = \sup_{c \in C} F(c) \nsim \chi^2$$

The asymptotic distribution of F_s is non-standard.

Hansen (1996) shows that the asymptotic distribution may be approximated by a bootstrap procedure in general, and Hansen (1997) gives the analytic form of the asymptotic distribution for testing against 2-regime SETAR model.

3.2 Test for STAR nonlinearity

For STAR model, it is complicated and dependent on the number of regimes. We consider a 2-regime STAR model with $z_t = y_{t-1}$ as follows

$$y_t = \phi_1' x_t (1 - G(y_{t-1}; \gamma, c)) + \phi_2' x_t G(y_{t-1}; \gamma, c) + \varepsilon_t,$$
$$G(y_{t-1}; \gamma, c) = \frac{1}{1 + \exp(-\gamma[y_{t-1} - c])},$$

- The null hypothesis (linearity): $H_0 : \phi_1 = \phi_2 \Leftrightarrow H_0' : \gamma = 0$.
  - If $\gamma = 0$, $G(y_{t-1}; 0, c) = 0.5$ for all values of y_{t-1} and the STAR model reduces to an AR model with parameters $(\phi_1 + \phi_2)/2$.
  - Whichever formulation of the null hypothesis is used, the model contains unidentified parameters (e.g. c).

- Consider again the STAR model and rewrite this as

$$y_t = \frac{1}{2}(\phi_1 + \phi_2)' x_t + (\phi_2 - \phi_1)' x_t G^*(y_{t-1}; \gamma, c) + \varepsilon_t, \tag{39}$$

where

$$G^*(y_{t-1}; \gamma, c) = G(y_{t-1}; \gamma, c) - 1/2.$$

Notice that under the null hypothesis $\gamma = 0$, $G^*(y_{t-1}; 0, c) = 0$.

- Luukkonen, Saikkonen and Teräsvirta (1988) suggest approximating the function $G^*(y_{t-1}; \gamma, c)$ with a first-order Taylor approximation around $\gamma = 0$, that is,

$$\begin{aligned} T_1(y_{t-1}; \gamma, c) &\approx G^*(y_{t-1}, 0, c) + \gamma \left. \frac{\partial G^*(y_{t-1}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} \\ &= \frac{1}{4}\gamma(y_{t-1} - c). \end{aligned} \tag{40}$$

- After substituting $T_1(\cdot)$ for $G^*(\cdot)$ in (39) and rearranging terms

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \beta_1'\tilde{x}_t y_{t-1} + \eta_t, \tag{41}$$

where $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$ and $\beta_j' = (\beta_{1,j}', \ldots, \beta_{p,j}')'$, $j = 0, 1$. The relationships between the parameters in the auxiliary regression model (41) and the parameters in the STAR model (39) can be shown to be

$$\beta_{0,0} = (\phi_{0,1} + \phi_{0,2})/2 - \frac{1}{4}\gamma c(\phi_{0,2} - \phi_{0,1}), \tag{42}$$

$$\beta_{1,0} = (\phi_{1,1} + \phi_{1,2})/2 - \frac{1}{4}\gamma(c(\phi_{1,2} - \phi_{1,1}) - (\phi_{0,2} - \phi_{0,1})), \tag{43}$$

$$\beta_{i,0} = (\phi_{i,1} + \phi_{i,2})/2 - \frac{1}{4}\gamma c(\phi_{i,2} - \phi_{i,1}), \quad i = 2, \ldots, p, \tag{44}$$

$$\beta_{i,1} = \frac{1}{4}\gamma c(\phi_{i,2} - \phi_{i,1}), \quad i = 1, \ldots, p. \tag{45}$$

- The above equations demonstrate that the restriction $\gamma = 0$ implies $\beta'_{i,1} = 0$ for $i = 1, \ldots, p$. Hence testing the null hypothesis $H'_0 : \gamma = 0$ in (39) is equivalent to testing the null hypothesis $H''_0 : \beta'_1 = 0$ in (41)

$$H'_0 : \gamma = 0 \quad \Leftrightarrow \quad H''_0 : \beta'_1 = 0.$$

- This null hypothesis can be tested by a standard variable addition test-statistic in a straightforward manner. The test is usually referred to as an LM-type-statistic, given by

$$LM_1 = \frac{n(SSR_0 - SSR_1)}{SSR_1} \sim \chi^2(p)$$

Under the null hypothesis of linearity, the test-statistic has a χ^2 distribution with p degrees of freedom asymptotically.

The above test statistic does not have power in situations where only the intercept is different across regimes – that is, when $\phi_{0,1} \neq \phi_{0,2}$ but $\phi_{i,1} = \phi_{i,2}$ for $i = 1, \ldots, p$.

- Luukkonen, Saikkonen and Teräsvirta (1988) suggest remedying this deficiency by replacing the transition function $G^*(y_{t-1}; \gamma, c)$ by a third-order Taylor approximation instead, that is,

$$\begin{aligned} T_3(y_{t-1}; \gamma, c) &\approx \gamma \left. \frac{\partial G^*(y_{t-1}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} \\ &\quad + \frac{1}{6}\gamma^3 \left. \frac{\partial^3 G^*(y_{t-1}; \gamma, c)}{\partial \gamma^3} \right|_{\gamma=0} \\ &= \frac{1}{4}\gamma(y_{t-1} - c) + \frac{1}{48}\gamma^3 (y_{t-1} - c)^3, \end{aligned} \tag{46}$$

where the second derivative of G^* evaluated at $\gamma = 0$ equals zero.

- Using this approximation yields the auxiliary model

$$y_t = \beta_{0,0} + \beta_0' \tilde{x}_t + \beta_1' \tilde{x}_t y_{t-1} + \beta_2' \tilde{x}_t y_{t-1}^2 + \beta_3' \tilde{x}_t y_{t-1}^3 + \eta_t, \tag{47}$$

  where $\beta_{0,0}'$ and the β_j', $j = 1, 2, 3$, again are functions of the parameters ϕ_1, ϕ_2, γ and c.

- Inspection of the exact relationships demonstrates that the null hypothesis $H_0' : \gamma = 0$ now corresponds to $H_0'' : \beta_1' = \beta_2' = \beta_3' = 0$, which again can be tested by a standard LM-type test.

- Under the null hypothesis of linearity, the test-statistic has a χ^2 distribution with $3p$ degrees of freedom asymptotically.

In small samples, the usual recommendation is to use F-versions of the LM test-statistics, as these have better size and power properties. The F-version of the test-statistic based on (47) can be computed as follows:

1. estimate the model under the null hypothesis of linearity by regressing y_t on x_t. Compute the residuals $\tilde{\varepsilon}_t$ and the sum of squared residuals $SSR_0 = \sum_{t=1}^{n} \tilde{\varepsilon}_t^2$;
2. estimate the auxiliary regression of $\tilde{\varepsilon}_t$ on x_t and $\tilde{x}_t y_{t-1}^j$, $j = 1, 2, 3$, and compute the sum of squared residuals from this regression SSR_1;
3. the F-version of LM test-statistic can be computed as

$$LM_F = \frac{(SSR_0 - SSR_1)/3p}{SSR_1/(n - 4p - 1)} \sim F(3p, n - 4p - 1), \qquad (48)$$

   and is approximately F-distributed with $3p$ and $n - 4p - 1$ degrees of freedom under the null hypothesis.

Choosing the transition variable

Teräsvirta (1994) suggests that the LM-type test (48) can also be used to select the appropriate transition variable in the STAR model.

The statistic is computed for several candidate transition variables and the one for which the p-value of the test is smallest is selected as the true transition variable.

The rationale behind this procedure is that the test should have maximum power if the alternative model is correctly specified – that is, if the correct transition variable is used.

Simulation results in Teräsvirta (1994) suggest that this approach works quite well, at least in a univariate setting.

3.3 Test for MS-type nonlinearity

When assessing the relevance of the MSAR model, a natural approach is to use a Likelihood Ratio (LR)-statistic, which tests the null hypothesis of linearity against the alternative of a MSAR model – that is, $H_0 : \phi_1 = \phi_0$ is tested by means of the test-statistic

$$LR = \mathcal{L}_{MS} - \mathcal{L}_{AR}, \tag{49}$$

where $\mathcal{L}_{MS}$ and $\mathcal{L}_{AR}$ are the values of the log likelihood functions corresponding to the MSAR and AR models, respectively.

Note that

- Transition probabilities, p_{11} and p_{00}, are unidentified under the null since any value between 0 and 1 leaves the likelihood function unchanged.

- Moreover, the scores are identically zero when evaluated at the null hypothesis.

Under these conditions the asymptotic null distribution of the likelihood ratio test statistic is not a χ^2 with the corresponding degrees of freedom and it has a nonstandard distribution which cannot be characterized analytically (Hansen, 1992).

Existing testing methods

- Hansen (1992, 1996) proposes a theory of hypothesis testing that avoids the above problems, but that procedure only gives bounds for the LR statistic and requires extensive computation, which involves large scale simulation and optimization over a three-dimensional grid.

- Carrasco et al. (2004) propose a test for the stability of parameters in MS models, which only requires the estimation of the model under the null. This is a great advantage over the tests of Hansen (1992, 1996), given that it is well-known that the estimation of MS models is particularly burdensome.

- Recently, Di Sanzo (2009, Testing for linearity in Markov switching models: a bootstrap approach.

The bootstrap algorithm runs as follows:

S1: Construct some estimation of the coefficients of the model under the null hypothesis of linearity, $\hat{\theta}_0 = \{\hat{\phi}, \hat{\sigma}\}$. For example, using the MLE.

S2: Compute the standardized residuals under H_0 as follows:

$$\hat{u}_t = (y_t - x_t'\phi)/\hat{\sigma}, \quad t = p+1, \ldots, T.$$

S3: Estimate the model under H_1 and compute the LR statistic

$$LR = 2[L(\hat{\theta}|\mathcal{F}_T) - L(\hat{\theta}_0|\mathcal{F}_T)],$$

where $\hat{\theta}$ denotes the unrestricted MLE of θ, and $\hat{\theta}_0$ is its MLE under H_0, which is computed in S1.

S4: Generate the bootstrap errors u_t^*, $t = p+1, \ldots, T$, by sampling with replacement from the set of standardized residual $\hat{u}_t$. The bootstrap sample is constructed as follows:

$$y_t^* = x_t' \hat{\phi} + \hat{\sigma} u_t^*.$$

The initial values for $(y_1, \ldots, y_p)$ can be fixed as the observed values.

S5: Use the bootstrapped series $\{y_t^*\}$ to calculate the LR statistic, denoted by LR^*.

S6: Repeat S4 $-$ S5 steps M times.

This bootstrap algorithm provides a distribution of LR^*, which is the bootstrap distribution of LR. The bootstrap p-value is then computed as

$$p_M = \frac{card(LR^* \geq LR)}{M},$$

which is the fraction of LR^* values which is greater than the observed value LR.