Nonlinear Models & Their Applications

Does nonlinearity exist in financial TS?

Yes, especially in volatility & high-freq data

We focus on simple nonlinear models & neural networks

What is a linear time series?

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where μ is a constant, ψ_i are real numbers with $\psi_0 = 1$, and $\{a_t\}$ is an iid $(0, \sigma_a^2)$.

General concept: Let F_h be info. available at h.

Conditional mean:

$$\mu_t = E(x_t | F_{t-1}) \equiv g(F_{t-1}),$$

Conditional variance:

$$\sigma_t^2 = \mathrm{Var}(x_t | F_{t-1}) \equiv h(F_{t-1})$$

where $g(.)$ and $h(.)$ are well-defined functions with $h(.) > 0$.

For a linear series, $g(.)$ is a linear function of F_{t-1} and $h(.) = \sigma_a^2$.

Statistics literature: focuses on $g(.)$

e.g. bilinear models of Granger and Andersen (1978), TAR model of Tong (1978, 1990), etc.

Econometrics literature: focuses on $h(.)$

Some specific models

TAR model: a piecewise linear model in the threshold space.

Example: 2-regime AR(1) model

$$x_t = \begin{cases} -1.5x_{t-1} + a_t & \text{if } x_{t-1} < 0, \\ 0.5x_{t-1} + a_t & \text{if } x_{t-1} \geq 0, \end{cases}$$

where a_t's are iid $N(0,1)$.

Here the delay is 1 time period, x_{t-1} is the **threshold** variable, and the threshold is 0. The threshold divides the threshold space into two regimes with Regime 1 denoting

$x_{t-1} < 0$.

What is so special about this model? See the time plot.

The model shows some special features: (a) asymmetry in rising and declining patterns, (more obs are positive than negative) (b) the mean of x_t is not zero even though there is no constant term in the model, (c) the lag-1 coefficient may be greater than 1 in absolute value.

Financial application:

Modeling asymmetry in volatility (recall EGARCH model)

Example: Daily log returns of IBM stock from July 3, 1962 to December 31, 1999 for 9442 observations. See the plot in the text (p. 132).

AR(2)-GARCH(1,1) model:

$$\begin{aligned} r_t &= 0.067 - 0.023 r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.031 + 0.076 a_{t-1}^2 + 0.915 \sigma_{t-1}^2 \end{aligned}$$

Std residuals: $Q(10) = 11.31(0.33)$ and $Q(20) = 27.00(0.14)$

Sq. std. res.: $Q(10) = 11.86(0.29)$ and $Q(20) = 19.19(0.51)$.

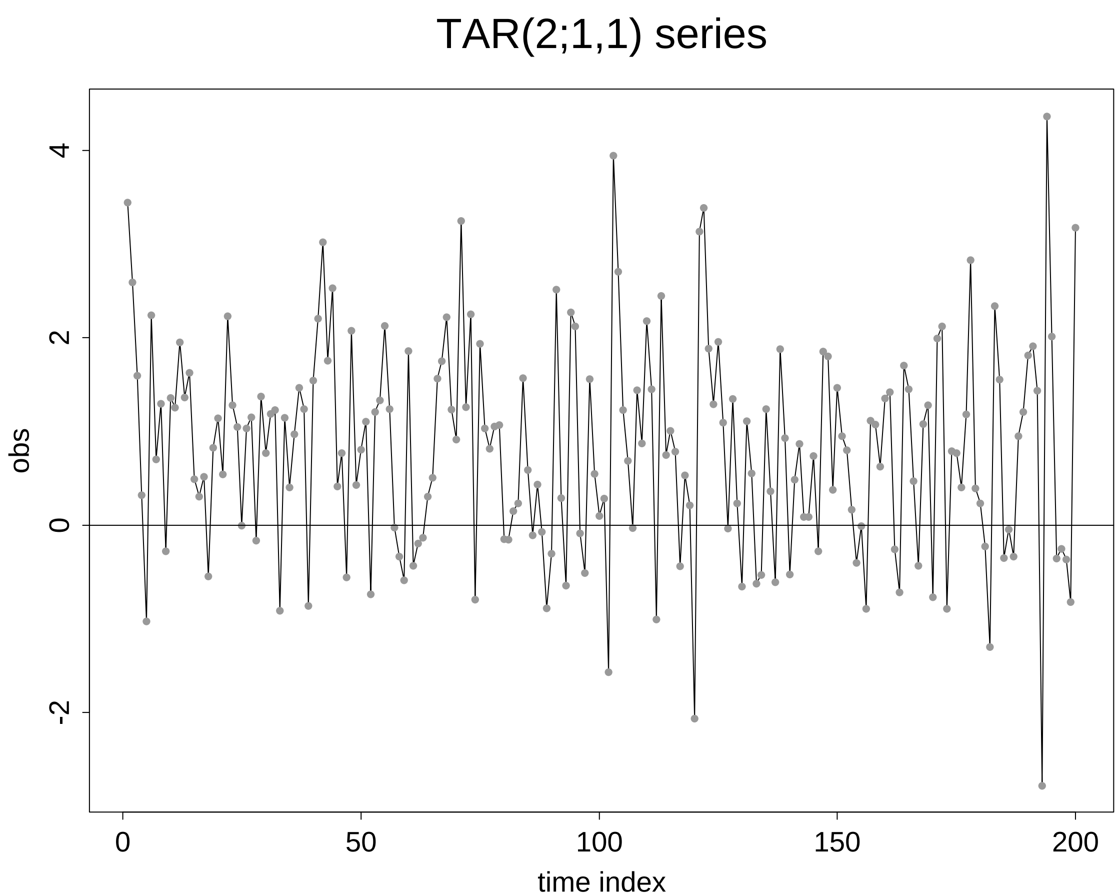


Figure 1: A simulated two-regime TAR process

AR(2)-TAR-GARCH(1,1) model

$$\begin{aligned} r_t &= 0.043 - .022 r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.098 a_{t-1}^2 + 0.954 \sigma_{t-1}^2 \\ &+ (0.060 - .052 a_{t-1}^2 - 0.069 \sigma_{t-1}^2) I(a_{t-1} > 0). \end{aligned}$$

Further simplication:

$$\begin{aligned} r_t &= 0.068 - .027 r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= (1.0 - 0.926) a_{t-1}^2 + 0.926 \sigma_{t-1}^2 \\ &+ (0.045 - .038 a_{t-1}^2 + 0.021 \sigma_{t-1}^2) I(a_{t-1} > 0). \end{aligned}$$

Rewrite the TAR-GARCH(1,1) as

$$\sigma_t^2 = \begin{cases} 0.074 a_{t-1}^2 + 0.926 \sigma_{t-1}^2 & \text{if } a_{t-1} \le 0 \\ 0.045 + 0.036 a_{t-1}^2 + 0.947 \sigma_{t-1}^2 & \text{if } a_{t-1} > 0, \end{cases}$$

Discussion: The asymmetry in volatility is clearly seen. When $a_{t-1} < 0$, the volatility follows an IGARCH model without any drift. However, the volatility follows a GARCH(1,1) model when a_{t-1} is positive. But the persistent coefficient is close to unity for $a_{t-1} > 0$.

STAR model

A 2-regime star model:

$$x_t = c_0 + \sum_{i=1}^{p} \phi_{0,i} x_{t-i} + F(\frac{x_{t-d} - \Delta}{s})(c_1 + \sum_{i=1}^{p} \phi_{1,i} x_{t-i}) + a_t$$

Discussion: A stochastic mixture of two linear AR models.

Example: Monthly simple stock returns for 3M stock from February 1946 to December 1997.

An ARCH(2) model:

$$R_t = 0.014 + a_t, \quad a_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = 0.003 + 0.108 a_{t-1}^2 + 0.151 a_{t-2}^2$$

A star model:

$$R_t = 0.017 + a_t, \quad a_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = (.002 + .256 a_{t-1}^2 + .141 a_{t-2}^2)$$

$$+ \frac{1}{1 + \exp(-1000 a_{t-1})} (.002 - .314 a_{t-1}^2)$$

Implication:

For a large negative a_{t-1},

$$\sigma_t^2 = 0.002 + 0.256 a_{t-1}^2 + 0.141 a_{t-2}^2.$$

For a large positive a_{t-1},

$$\sigma_t^2 = 0.005 - 0.058 a_{t-1}^2 + 0.141 a_{t-2}^2.$$

Further simplification: To ensure no negative coefficient for a_{t-1}^2, we may require that the two a_{t-1}^2 coefficients sum to zero. This results in the model

$$R_t = 0.017 + a_t, \quad a_t = \sigma_t \epsilon_t,$$

$$\begin{aligned} \sigma_t^2 &= (.0024 + .261 a_{t-1}^2 + .132 a_{t-2}^2) \\ &+ \frac{1}{1 + \exp(-1000 a_{t-1})} (.0015 - .261 a_{t-1}^2) \end{aligned}$$

Markov switching model

Two-state MS model:

$$x_t = \begin{cases} c_1 + \Sigma_{i=1}^p \phi_{1,i} x_{t-i} + a_{1t} & \text{if } s_t = 1, \\ \\ c_2 + \Sigma_{i=1}^p \phi_{2,i} x_{t-i} + a_{2t} & \text{if } s_t = 2, \end{cases}$$

where s_t assumes values in $\{1,2\}$ and is a first-order Markov chain with trans. prob.

$$P(s_t = 2|s_{t-1} = 1) = w_1, \quad P(s_t = 1|s_{t-1} = 2) = w_2,$$

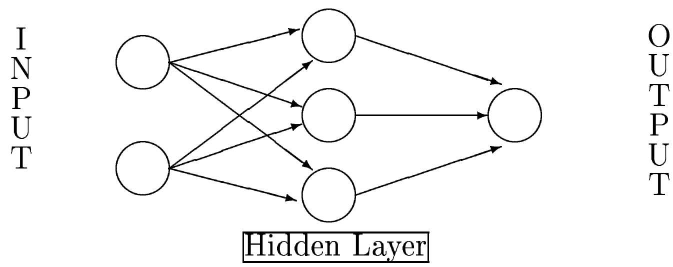
where $0 \leq w_1 \leq 1$ is the probability of switching out State 1 from time $t-1$ to time t. A large w_1 means that it is easy to switch out State 1, i.e. cannot stay in State 1 for long. The inverse, $1/w_1$, is the expected duration (number of time periods) to stay in State 1. Similar idea applies to w_2.

Example: Growth rate of US quarterly real GNP 47-91. See the plot in the text (p.137).

State 1							
Par	c_i	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ_i	w_i
Est	0.909	0.265	0.029	−0.126	−0.110	0.816	0.118
S.E	0.202	0.113	0.126	0.103	0.109	0.125	0.053
State 2							
Est	−0.420	0.216	0.628	−0.073	−0.097	1.017	0.286
S.E	0.324	0.347	0.377	0.364	0.404	0.293	0.064

Discussion

- Regime 2, which has a negative expectation, denotes "recession" periods. The S.E. of the estimates are large due to the small number of data in the regime.
- The expected durations for Regime 1 and 2 are 8.5 and 3.5 quarters, respectively. $(1/w_i)$



Neural networks

- a semi-parametric approach to data analysis
- Structure of a network
  - Output layer
  - Input layer
  - Hidden layer
  - Nodes
- Activation function:
  - Logistic function:

$$\ell(z) = \frac{\exp(z)}{1 + \exp(z)}$$

– Heaviside (or threshold) function:

$$H(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

- Use $\ell(z)$ for the hidden layer

Feed-forward neural network:

Hidden node:

$$x_j = f_j(\alpha_j + \sum_{i \to j} w_{ij} x_i)$$

where $f_j(.)$ is an activation function which is typically taken to be the logistic function

$$f_j(z) = \frac{\exp(z)}{1 + \exp(z)},$$

α_j is called the bias, the summation $i \to j$ means summing over all input nodes feeding to j, and w_{ij} are the weights.

Output node:

$$y = f_o(\alpha_o + \sum_{j \to o} w_{jo} x_j),$$

where the activation function $f_o(.)$ is either linear or a Heaviside function. By a Heaviside function, we mean $f_o(z) = 1$ if $z > 0$ and $f_o(z) = 0$, otherwise.

General form:

$$y = f_o\left[\alpha_o + \sum_{j \to o} w_{jo} f_j\left(\alpha_j + \sum_{i \to j} w_{ij} x_i\right)\right].$$

With direct connections from the input layer to the output layer:

$$y = f_o\left[\alpha_o + \sum_{i \to o} w_{io} x_i + \sum_{j \to o} w_{jo} f_j\left(\alpha_j + \sum_{i \to j} w_{ij} x_i\right)\right],$$

<u>Training and forecasting</u>

Divide the data into training and forecasting subsamples.

Training: build a few network systems

Forecasting: based on the accuracy ofout-of-sample forecasts to select the "best" network.

Example: Monthly log returns of IBM stock 26-99.

See text for details.

Some references

Related to Credit Risk

- Elmer & Borowski (1988, *Financial Management*): bankruptcy prediction
- Messier & Hansen (1988, *Management Science*): forecasting business failures

Related to Bond Rating: two articles in the book *Neural Networks in Capital Markets*, ed. A.P. Refens, Wiley, 1994.

- Moody & Utans
- Singleton & Surkan

Exercise: use S-plus to gain some experience.