Specialist Courses

Daniele Zago

January 13, 2022

Contents

Lecture 6: Nonlinear time series models

2022-01-11

A lot of research in the last 30 years has been spent to develop time-series models which can detect nonlinear patterns in the data. We refer again to Wold's theorem (**??**) and consider a time series of the form

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + V_t,$$

where $\psi_0 = 1$ and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$, $a_t \sim \text{WN}(0, \sigma^2)$ and a_j's are uncorrelated.

Def. (General linear process)
X_t is said to be ***linear*** if if can be written as

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

such that $\sum |\psi_j| < \infty$ and $a_j \overset{\text{iid}}{\sim} \text{WN}(0, \sigma^2)$. The above process is sometimes called ***general linear process***.

Limitations Linear models are limited in the sense that they cannot model ***strong asymmetries*** in data, ***irregular jumps***, and ***switching regimes***.

6.1 Nonlinear framework

Whereas linearity is well-defined, non-linearity is hardly definite and the early development of nonlinear time series focused on various parametric forms: ARCH, GARCH, threshold models, ...

In this lectures we emphasize simple parametric models which are applicable without overly-complex specifications. There are examples of *explicit* and *implicit* approaches, which differ in the way they are represented:

- *Implicit*: ARMA model with non-gaussian innovations.
- *Explicit*: $X_t = h(X_{t-1}, \ldots, X_{t-k})$.

Explicit models have surpassed implicit modelling since it is in general difficult to identify the correct distribution of the white-noise terms.

Attention In the nonlinear setting, the tools of standard analysis (ACF, PCF, ...) are not helpful since they only detect linear patterns.

Def. (Nonlinear model)
The genreal representation is

$$X_t = f(a_t, a_{t-1}, a_{t-2}),$$

where $a_t \overset{\text{iid}}{\sim} \text{WN}(0, \sigma^2)$ and $f(\cdot)$ is some nonlinear function.

Remark We could linearize the above model by considering the Taylor series around zero (Volterra series)

$$X_t = \mu + \sum_{i,j} b_{ij} a_{i,t-j} + \sum_{ijkl} b_{ij,kl} a_{i,t-j} a_{k,t-l} + \dots \tag{1}$$

The Volterra series (1) is usually too complicated unless severely truncated.

Other very general models are ARMA models with time-dependent parameters.

The model for nonlinear time series can be written in terms of the conditional mean and variance,

$$\mu_t = \mathbb{E}[X_t|\mathcal{F}_{t-1}] = g(F_{t-1})$$

$$\sigma_t^2 = \mathbb{V}[X_t|\mathcal{F}_{t-1}] = h(F_{t-1})$$

where g and h are well-defined nonlinear function with $h(\cdot) > 0$. If $g(\cdot)$ is nonlinear and $h(\cdot)$ is constant, then X_t is nonlinear in mean, for example

$$X_t = \varepsilon_t + \alpha\varepsilon_{t-1}^2.$$

Otherwise, if $h(\cdot)$ is time-variant then X_t is nonlinear in variance. All GARCH models are of this type.

6.2 NLAR(1)

We consider the simplest conditional mean model given by

$$X_t = g(X_{t-1}, \vartheta) + a_t,$$

where ϑ is a vector of parameters and $a_t \sim$ IID. It's natural to consider functions which are nearly linear as first candidates, but also more extreme nonlinear functions if needed.

There are few papers on nonlinear autoregressive processes, since the fact that there are too many nonlinear functions that we can consider renders this class of processes unusable for statistical analysis.

6.3 Reversibility

Def. (Reversibility)
The stationary sequence X_t is ***time-reversible*** if the finite-dimensional distributions of $(X_1, X_2, \dots, X_n)$ and $(X_n, \dots, X_2, X_1)$ is the same for any n.

Usage Since i.i.d sequences and ARMA models are time-reversible, we can use time-reversibility to detect deviations from the Gaussianity-linearity hypothesis. For example, Chen et al (2000) look at the test statistic

$$\mathbb{E}[\sin(\omega(X_t - X_{t-k}))],$$

and if the value is zero then the process is time-reversible. Other tests rely for example on differences between backward and forward autocorrelation.

6.4 Threshold AR models

Consider the change-point model

$$X_t = \begin{cases} \varphi_1 X_{t-1} + a_t & \text{if } X_{t-1} < r \\ \varphi_2 X_{t-1} + a_t & \text{if } X_{t-1} \geq r \end{cases}$$

we call this model the *self-exciting* threshold-AR model. If X_{t-1} is replaced by an exogenous variable Z_{t-d}, then this model is called *threshold-autoregressive* (TAR). Using piecewise linear models we can obtain a better approximation of the conditional mean equation.

Def. (SETAR model)
We define the ***self-exciting TAR model*** (***SETAR***) with threshold X_{t-d} if

$$X_t = \varphi_0^{(j)} + \varphi_1^{(j)} X_{t-1} + \ldots + \varphi_p^{(j)} X_{t-p} + a_t^{(j)} \quad \text{if } \gamma_{j-1} \leq X_{t-d} \leq \gamma_j,$$

where $j = 1, \ldots, k$.

Remark In the above model, γ_j's are the thresholds and X_{t-d} is the threshold variable.

6.4.1 Estimation

Suppose that we have an observed time series $X_1, \ldots, X_n$ and we fix the order k of the SETAR model. We alternate a two-step procedure:

› First we assume that the partition A_j and orders p_j are known, so that we can use the least squares procedure to minimize the loss function

$$\mathcal{S}(\vartheta) = \sum_{j=1}^{k} S^{(j)} = \sum_{\substack{X_{t-d} \in A_j \\ p < t \leq n}} \left[X_t - (\varphi_0^{(j)} + \varphi_1^{(j)} X_{t-1} + \ldots + \varphi_p^{(j)} X_{t-p}) \right]^2$$

› We find the partition $\widehat{A}_j$ such that $\mathcal{S}(\widehat{\vartheta})$ is minimized.

Otherwise, to determine the autoregressive orders $p^{(j)}$'s we might use an information criteria such as BIC, AICC, ...

Testing for linearity becomes essential to check non-linearity when fitting nonlinear models. In general, there are

a) Tests for departure from linear models towards general nonlinear models (less powerful).

b) Tests for departure from linear models towards threshold autoregressive (more powerful).

6.4.2 Smooth Transition Autoregressive models

A generalization of the TAR model is the ***smooth transition autoregressive model*** (STAR) which allows for a smoother transition between the two regimes via a cumulative distribution function instead of a jump function.

Lecture 7: Markov-switching model

2022-01-13

A Markov-switching (MS) model changes the rules of the switching regimes from the threshold-autoregressive model, which are not deterministic anymore. A MS(p) model with two regimes can be defined as

$$X_t = \begin{cases} \alpha_1 + \sum_{i=1}^{p} \varphi_{1,i} X_{t-i} + a_{1,t} & \text{if } s_t = 1 \\ \alpha_2 + \sum_{i=1}^{p} \varphi_{2,i} X_{t-i} + a_{2,t} & \text{if } s_t = 2 \end{cases} \tag{2}$$

where $a_{i,t} \sim \text{IID}(0, \sigma_i^2)$. The state variable s_t is unobservable and we assume that it follows a first-order Markov chain with transition probabilities

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

and a particular choice of P drives the model behaviour. This is crucially different from the SETAR model, since the regimes are defined by the Markov chain and not simply determined by the past values of X_t. Thus, *forecast* of a Markov-switching model are linear combinations of forecasts produced by the sub-models.

Estimating a MS model is much harder since the states are not observable, therefore we need a ***filtering*** approach. The log-likelihood can be constructing recursively from some initial conditions since

$$f_{it} = f(X_t | s_t = i, X_{t-1}, \vartheta_i), \quad i = 1, 2,$$

and under normality this is a Gaussian density with parameters from Equation (2). Now, we can write the contribution by marginalizing s_t as

$$g(X_t | X_{t-1}, \vartheta_1, \vartheta_2) = f_{1t}\rho_{1t|t-1} + f_{2t}\rho_{2t|t-1}, \tag{3}$$

and Hamilton have shown that the optimal inference and forecast can be determined from the conditional likelihood in (3).

7.1 Bilinear models

This class of models is not very used/useful even though they are a natural extension of the ARMA model. A general bilinear model BL(p, q, r, s) can be written as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + a_t + \sum_{i=1}^{q} \vartheta_j a_{t-j} + \underbrace{\sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} a_{t-j}}_{\text{bilinear component}},$$

where $a_t \sim \text{IID}(0, \sigma_a^2)$. This model is too complex and therefore people usually study the Lower Triangular Bilinear Model

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + a_t + \sum_{i=1}^{q} \vartheta_j a_{t-j} + \underbrace{\sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i-j} a_{t-i}}_{\text{bilinear component}},$$

where X has only past values w.r. to a in the bilinear component.

› These models can model occasional outbursts in time series.

› BL model can have conditional heteroscedasticity, although GARCH-type models are better in this regard.

› ML procedures are used but asymptotic distribution is unknown.

› Probabilistic properties are often derived using the state-space representation.

7.2 Long-memory models

ARMA models are said to be short-term models, since their autocorrelation function usually tends to zero with an exponential decrease. Thus the $d \in \{0, 1\}$ parameter in an ARIMA(p, d, q) model controls the transition from short-memory to infinite memory.

Granger introduced an extension to the ARIMA model by considering $d \in [0, 1]$ yielding ARFIMA models. These models are necessary if we want to take into account series that show memory between $I(0)$ and $I(1)$ processes.

Properties ARFIMA models can take into account

› Presence of long-range cycles

› Slowly-decaying autocorrelation structures.

There are different definitions of long-memory processes:

1. In the time domain, a long-memory process is such that its autocorrelation function decays like a power function, i.e. if $\alpha \in (0, 1)$ and $c_\rho > 0$

$$\rho(k) = c_\rho k^{-\alpha}, \quad k \longrightarrow \infty.$$

2. In the frequency domain, a long-memory process is such that its spectral density is unbounded at zero, i.e.

$$f(\omega) \sim c_f \omega^{-\alpha}, \quad \omega \longrightarrow 0^+.$$

We can define the fractional difference operator using the Gamma function as

$$(1 - B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} B^j,$$

and when $d \in (0, 0.5)$ the ARFIMA(p, d, q) process is stationary with $\rho(k) \sim k^{2d-1}$. When $d \in (-0.5, 0)$ the process is stationary with intermediate memory, although in practice this is never used. For $d \in [0.5, 1)$ the process is mean-reverting even though it is not covariance-stationary.

In the following we will focus on ARFIMA(p, d, q) processes with $d \in (0, 0.5)$ which yields the most interesting type of process.

7.2.1 Estimation

Estimation approaches are mainly divided into two broad classes:

1. ML estimation, which requires specifying both p and q.
2. Semi-parametric or nonparametric approaches, where we assume that the ARMA component is relatively unimportant.

One of the best-known method is the semi-parametric GPH estimator introduced by Geweke and Porter-Hudack and developed by Robinson. This method approximates the spectral density near the origin,

$$f(\omega) \sim c_f \big(4 \sin^2(\omega/2)\big)^{-d},$$

and therefore we can apply the least-squares method to

$$\log l(\omega_j) = \log c_f - d \log \big(4 \sin^2(\omega_j/2)\big) + u_j,$$

where u_j are i.i.d error terms and ω_j are the Fourier frequencies. The problem with this method is its high variance in the estimates.

7.3 Integer autoregressive (INAR) models

Integer autoregressive (INAR) models can be used to model time series of counts, which are of particular interest in practice. In some cases, the discrete values can be approximated by Gaussian models, however for small values we need a more proper model.

Notation We introduce the ***thinning operator*** $\circ$ which substitutes the multiplication operator. Let $\alpha \in [0, 1]$, then we define

$$\alpha \circ X = \sum_{i=1}^{X} Y_i, \quad Y_i \text{ i.i.d r.v.'s with } \mathbb{E}[Y_i] = \alpha. \tag{4}$$

Typically, Y_i's are assumed to be i.i.d $\text{Ber}(\alpha)$, and therefore we have

$$Y_i \overset{\text{iid}}{\sim} \text{Ber}(\alpha) \implies \alpha \circ X | X \sim \text{Bin}(X, \alpha).$$

The INAR(1) model is defined as

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t,$$

where $\alpha \in [0, 1)$ and ε_t are i.i.d discrete random variables with mean $\mu_\varepsilon > 0$ and variance σ_ε^2. Usually we consider Poisson-distributed errors, but more flexible discrete distributions are possible.

The INAR(1) process is non-linear due to the thinning operator, but it's a member of the conditional linear first-order AR models,

$$\mathbb{E}[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon,$$

$$\mathbb{V}[X_t | X_{t-1}] = \alpha(1 - \alpha) X_{t-1} + \sigma_\varepsilon^2.$$

If $\varepsilon_t \sim \text{Pois}(\lambda)$, then $X_t \sim \text{Pois}\left(\frac{\lambda}{1-\alpha}\right)$.

Generalizing to the INAR(p) is not straightforward and depends on the definition of the thinning operator, e.g.

$$X_t = \alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p} + \varepsilon_t,$$

where the thinning operators are applied with independent Y_j's from (4).

Estimation is simple for Poisson innovations and $p = 1$, whereas for all other cases we have some problems. As for the forecast we use the median, since we want an integer value for our predictions and the mean is usually a real number.