- Threshold models constitute one of the earliest extensions of linear models to describe nonlinear dynamics [1].
- They are based on the fact that it is often possible to define different *states of the world* or *regimes*, and that it seems natural that the dynamics of the phenomenon under study at a given point in time should be dependent on the regime at this time.
- Examples: expansion and recession periods in economics or in an animal population.

- Usually, it is supposed that themodel describing the data in each regime is linear.
- The general form of a 2-regime threshold model is thus:

$$x_n = \begin{cases} a_{10} + a_{11} x_{n-1} + \cdots + a_{1p_1} x_{n-p_1} + \sigma_1 \varepsilon_n & \text{if } z_n \le r \\ a_{20} + a_{21} x_{n-1} + \cdots + a_{2p_2} x_{n-p_2} + \sigma_2 \varepsilon_n & \text{if } z_n > r \end{cases}$$

where z_n is the variable of interest defining the state at time *n*, and ε_n is an i.i.d. sequence.

- Some remarks:
  - The variable z_n may be endogenous. If $z_n = x_{n-d}$, for some lag value d, one speaks of a self-exciting threshold AR (SETAR) model.
  - The variable z_n may be exogenous, i.e. the state is defined by some other signal. When this signal is not observable, one usually refers to it as a switching model.
  - There is no a priori reason why the AR orders and innovation variances should be equal.

- This model is easily extended to a multiple regime one, conveniently described by:

$$x_n = \sum_{k=1}^{K} \left\{ a_{k0} + a_{k1} x_{n-1} + \cdots + a_{kp_k} x_{n-p_k} + \sigma_{\mathrm{k}} \varepsilon_n \right\} I\left(z_n \in A_k\right)$$

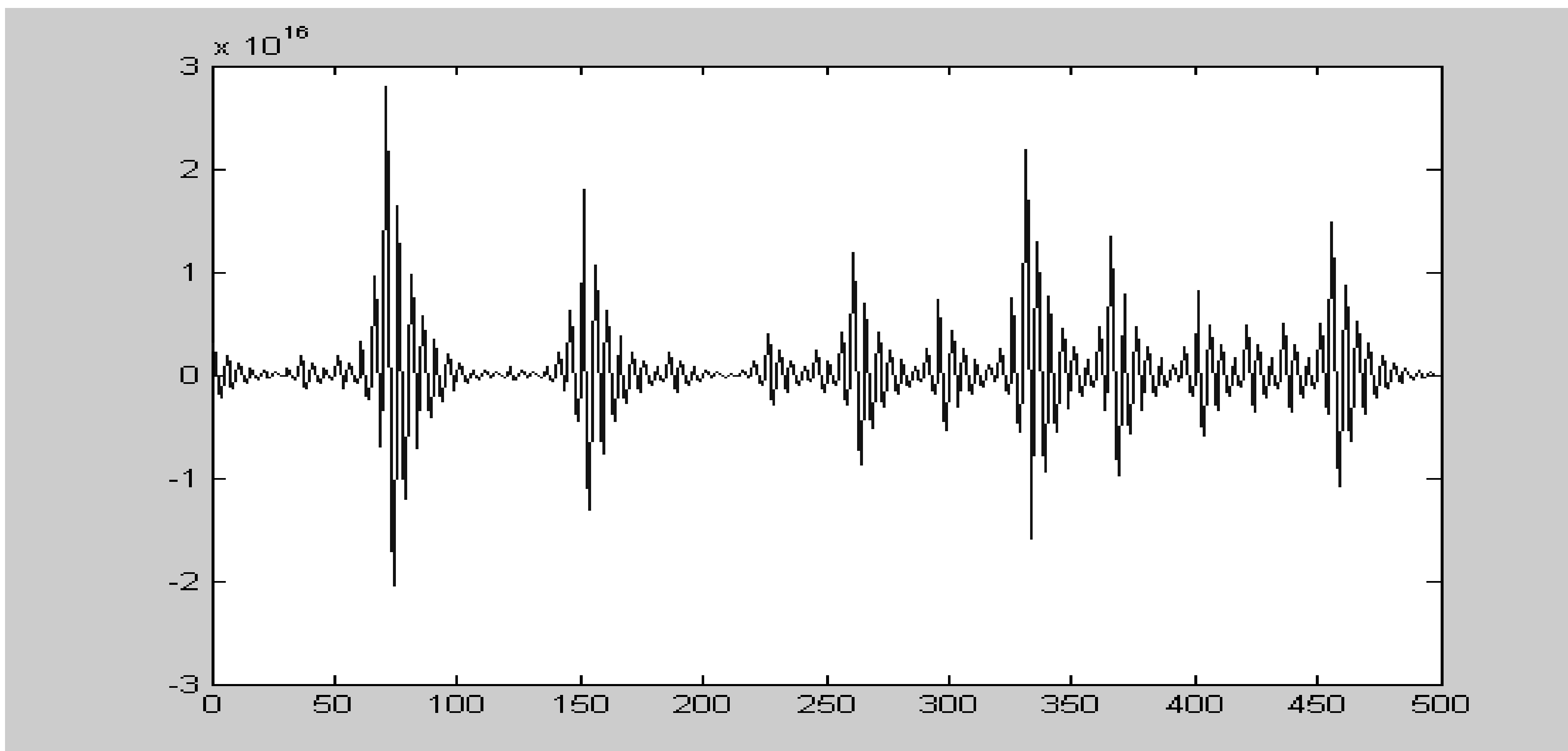
with $I(.)$ the indicator function and the subsets $\{A_k\}$ constitute a partition of the range of variation of z_n.

- We can use the sufficient condition (cf. *Basic Concepts*) for the ergodicity of a model, that expresses a mechanism of *drift back to the center*.
- In the case of a 2-regime model, a sufficient condition (which may be proved necessary) is that both AR sub-models are stable.
- In a multiple regime model, AR sub-models on the "borders" should be stable, while "sandwiched" ones can be unstable.

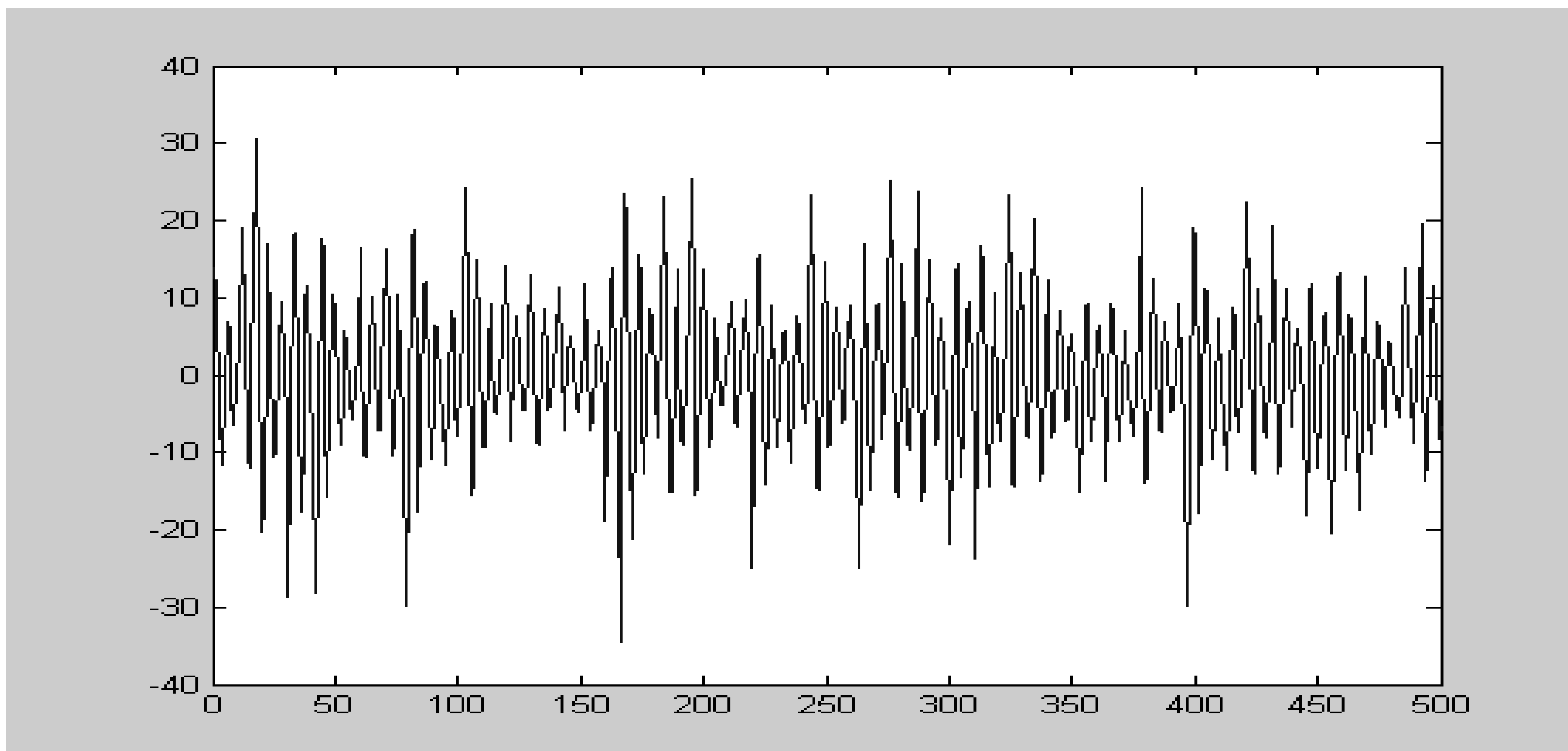
- Example: 3-regime model.

For $x_{n\text{-}1} \le -2$ or $x_{n\text{-}1} > 2$, stable AR(2), pole radius 0.9.

For $-2 < x_{n\text{-}1} \le 2$ unstable AR(2), pole radius 2.



- Note that if x_{n-2} is used instead of x_{n-1} as the threshold variable, the aspect of the signal changes notably.

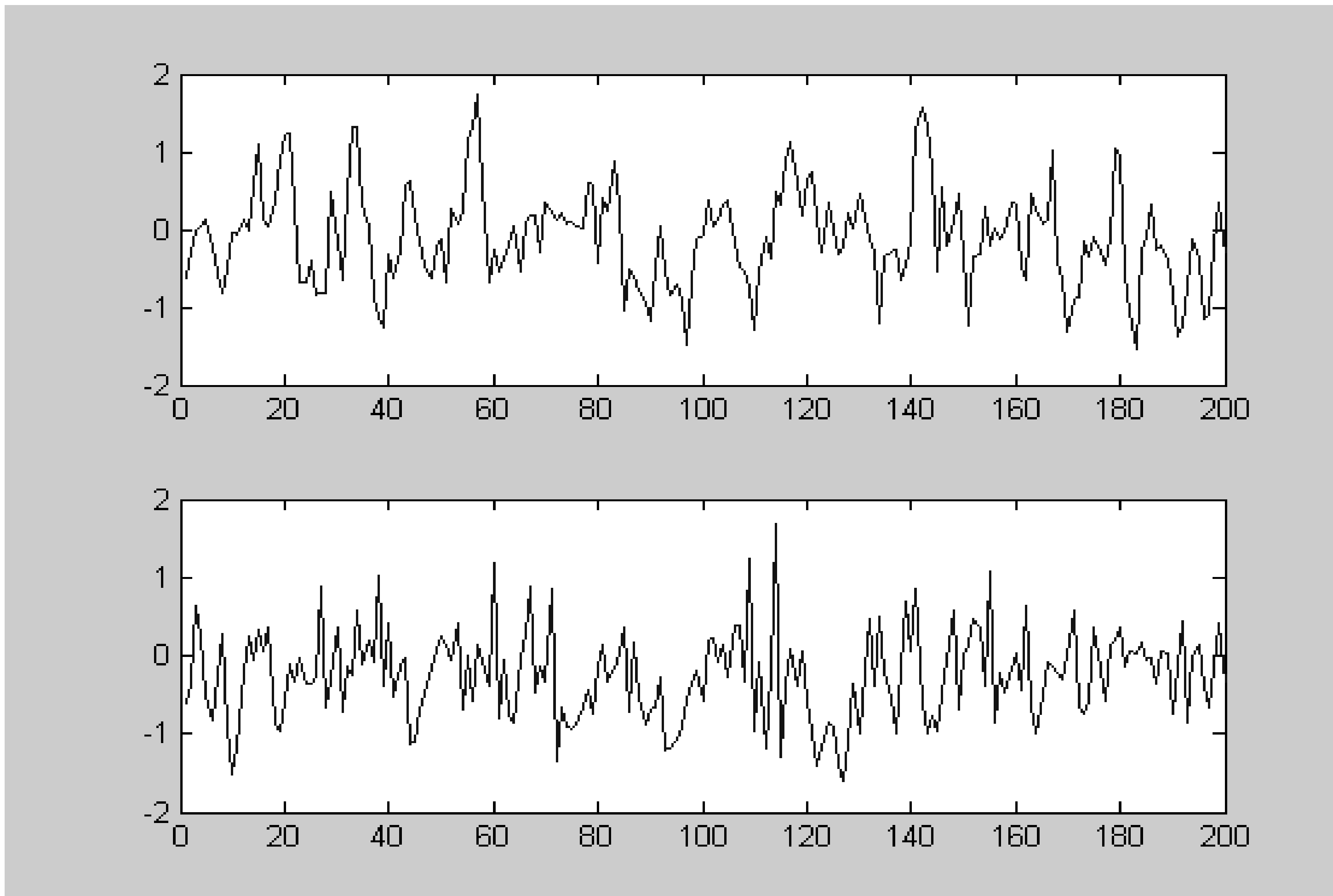


- When it comes to low-order SETAR models, scatter plots of the signal under study are often very insightful.
- Let us consider for instance the simple two-regime model:

$$x_n = \begin{cases} 0.7x_{n-1} + \varepsilon_n & \text{if } x_{n-1} \leq r \\ -0.7x_{n-1} + \varepsilon_n & \text{if } x_{n-1} > r \end{cases}$$

with ε_n an independent $N(0,0.25)$ sequence.

- Typical realizations:

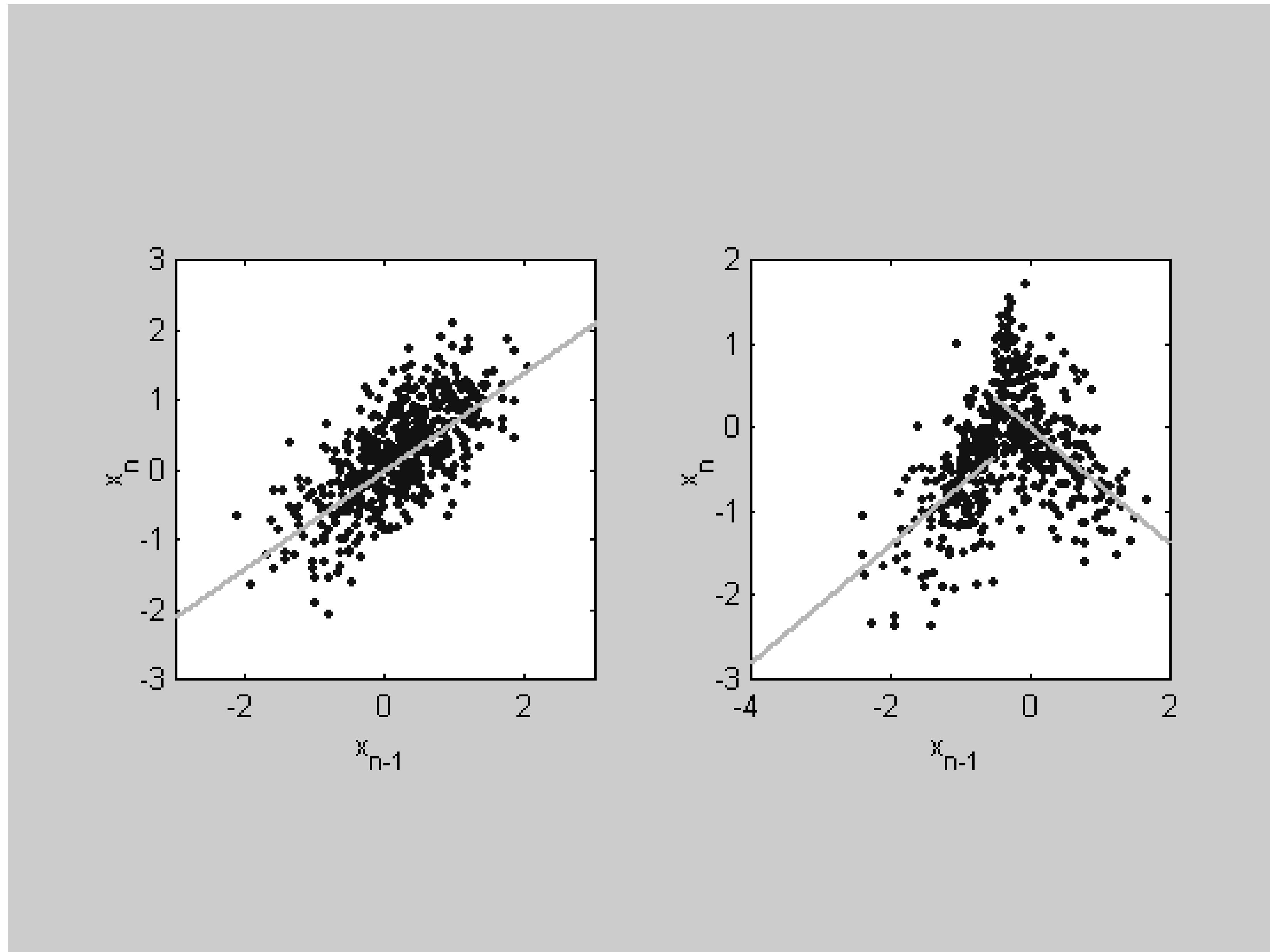


$r = -\infty$ (linear AR)

$r = -0.5$

- Corresponding scatter plots:

$r = -\infty$
(linear AR)



$r = -0.5$

- Suppose for the time being that the AR orders p_k, the lag d, and the partition $\{A_k\}$, are fixed.
- The least squares estimator for the AR coefficient vectors $\boldsymbol{a}_k = [a_{k0}, \ldots, a_{kpi}]^{\mathsf{T}}$ is simply:

with:

$$\sum_{k=1}^{K} L(\mathbf{a}_k; d; A_k)$$

$$L(\mathbf{a}_k; d; A_k) = \sum_{x_{n-d} \in A_k} (x_n - a_{k0} - a_{k1} x_{n-1} - \cdots - a_{kp_k} x_{n-p_k})^2$$

i.e. each of the sub-models is estimated separately.

- Now the variances can be estimated as:

$$\hat{\sigma}_k^2 = \frac{1}{N_k} L(\mathbf{a}_k; d; A_k)$$

where N_k is the number of samples x_{n-d} in A_k.

- In conventional NAR models, if the are supposed to be Gaussian, then least squares estimation is close to maximum likelihood estimation. It is not true for SETAR models.

- It is of course due to the presence of multiple variances. The maximum likelihood estimate for Gaussian innovation is obtained by minimizing:

$$-\frac{1}{2}\sum_{k=1}^{K} L(\mathbf{a}_k; d; A_k) \big/ \sigma_k^2 - \frac{1}{2}\sum_{k=1}^{K} N_k \ln(\sigma_k)$$

- But if the variances are not too far apart both estimates will be quite close.

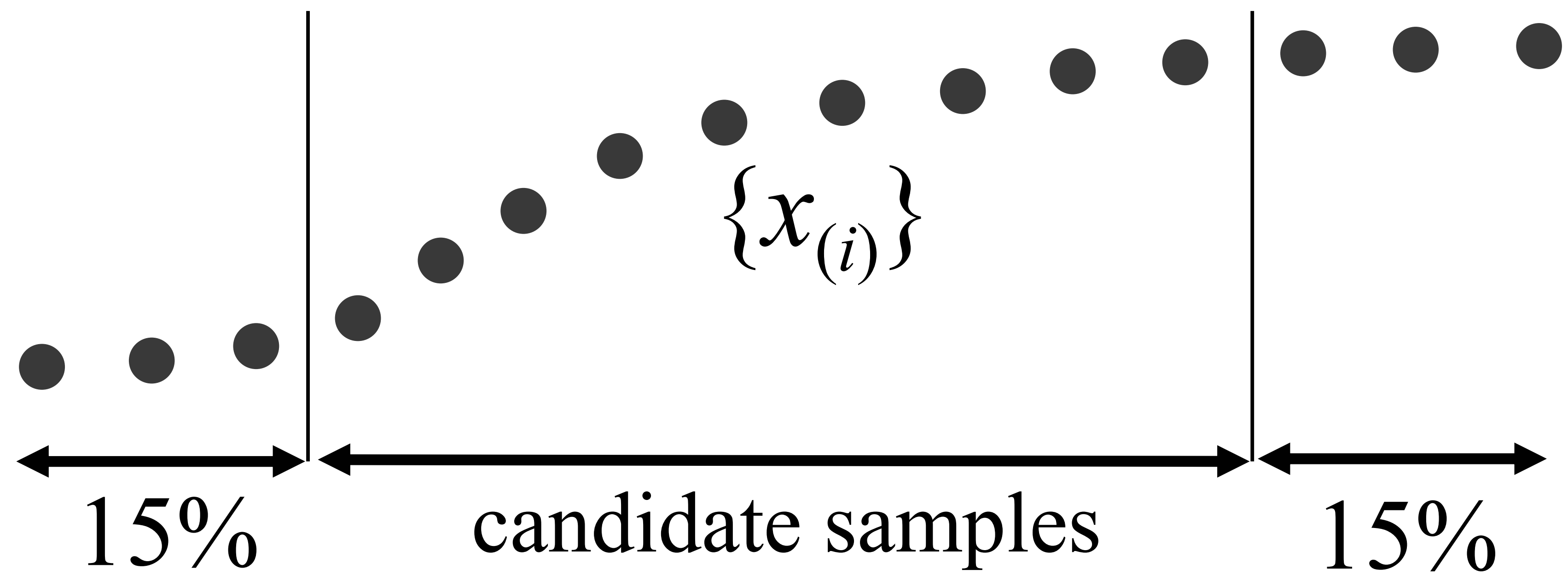
- Suppose now that only the AR orders p_k, the lag d are fixed. How can the partition be defined?
- The discussion will be limited to the 2-regime case (i.e. the threshold r must be defined), but is easily generalized to the multiple regime one.
- Since r is a real number, it could take any value in the range $[\min(x_n)\ \max(x_n)]$. But it is to be noted that *the least squares estimate of the model will change only when r crosses a sample value.*

- Also, the numbers of samples N_1 and N_2 that are involved in the least squares estimation of each sub-model should be large enough for the estimates to be reliable. A safe choice is that both N_1 and N_2 should be at least 15% of $N_1 + N_2$.
- So the idea is to sort the signal samples at hand in ascending order:

$$\{x_n\} \longrightarrow \{x_{(i)}\}$$

and determine the possible values for threshold r.

- Graphically:



- Then all that remains is to estimate the SETAR model for all possible threshold values, and select the one with smallest least squares error.

- Suppose now the lag d for the threshold variable is not known. One can fix a maximum value $d_{\max}$ (usually $\max[p_k]$) and try all values of d between 1 and $d_{\max}$ for all candidate thresholds.

- The last point is how to select the AR orders p_k. It is obviously possible to apply a model selection criterion such as MDL.

- The problem is that using a classical MDL formulation (written here for a 2-regime model):

$$\mathrm{MDL}(p_1, p_2) = N \ln(\sigma^2) + (p_1 + p_2 + 2) \ln(N)$$

with N the total number of samples and σ^2 the global error variance, typically penalizes SETAR models to much with respect to linear AR ones, especially if only a limited number of samples corresponds to one of the sub-models.

- This is why a special form of the MDL (similar formulation for other criteria) has been proposed:

$$\mathrm{MDL}(p_1, p_2) = N_1 \ln(\sigma_1^2) + N_2 \ln(\sigma_2^2) + (p_1 + 1)\ln(N_1) + (p_2 + 1)\ln(N_2)$$

i.e. the coding cost is considered separately for each sub-model and the corresponding samples.

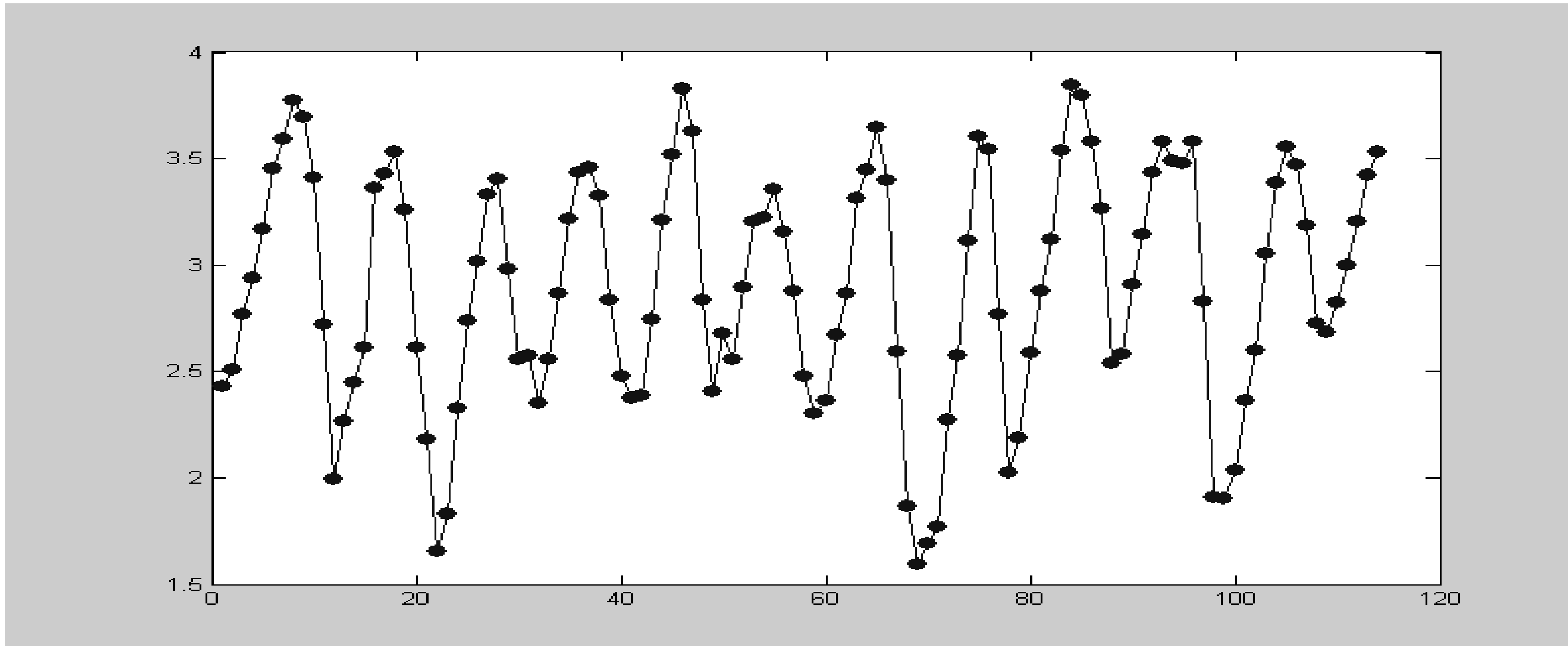
- The coding cost not taken into account is that of the sub-model number for each residual.

- To sum up, complete selection of a 2-regime SETAR model implies testing models for each possible pair of AR orders, all candidate thresholds and all lag values for the threshold variable.
- This is why having some a priori information (such as that given by scatter plots) may be worthwhile.

EXAMPLE OF SETAR MODELING (1)

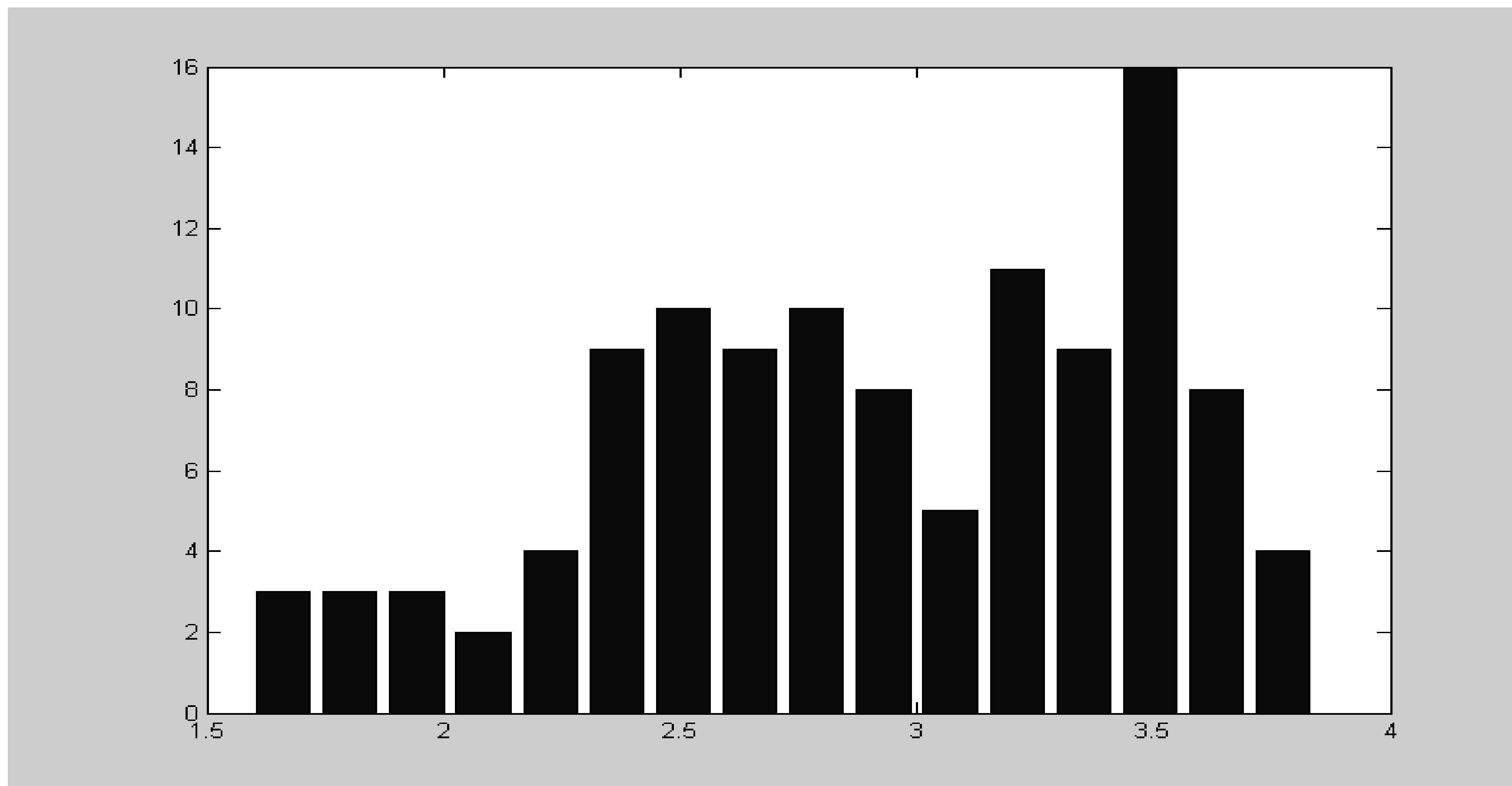
- The time series modeled is the (benchmark) *lynx* time series, more precisely its (base 10) logarithm.
- It corresponds to the number of lynx trapped in the Mackenzie River district of northwest Canada.
- It was early recognized by Moran, who first fitted a linear AR(2) model to this time series, that it presented nonlinear features.

EXAMPLE OF SETAR MODELING (2)



- It may be observed that the series is not time reversible. The phases of increase are typically slower than the phases of decrease.

- The histogram of the data also indicates a non Gaussian characteristic:

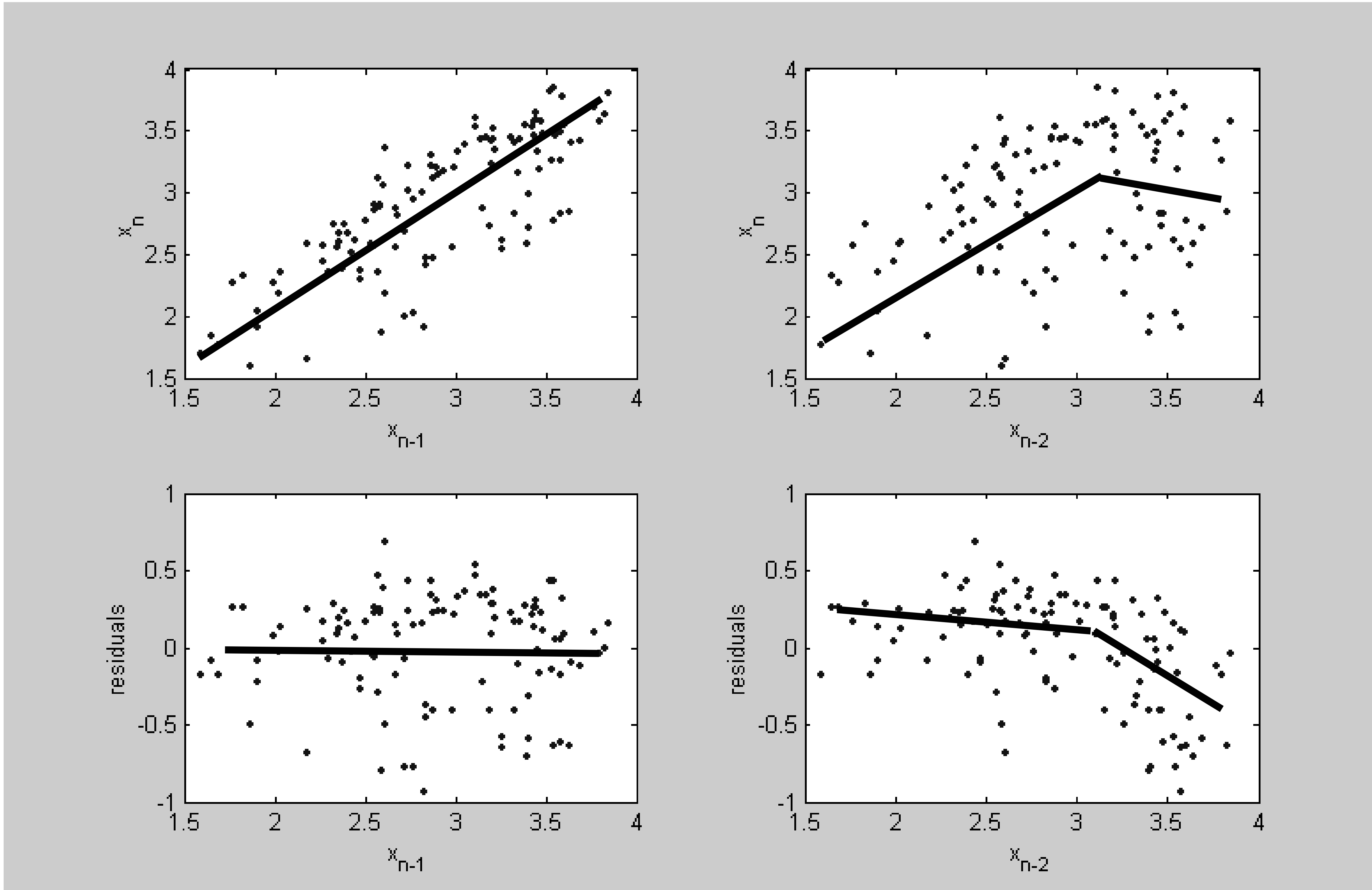


- It is instructive to visualize some scatter plots, namely x_n vs. x_{n-1}, and x_n vs. x_{n-2}. One may distinguish a "break" in the second plot.
- Of interest also are the plots of the residuals of a linear fit of x_n with respect to x_{n-1}, i.e. an estimate of $\mathsf{E}[x_n|x_{n-1}]$, which corresponds to:

$$\hat{x}_n = 0.620 + 0.788 x_{n-1}$$

vs. x_{n-1}, and x_{n-2}. Here also, a break is visible on the second plot.

EXAMPLE OF SETAR MODELING (5)



- Tong proposed the following SETAR model:

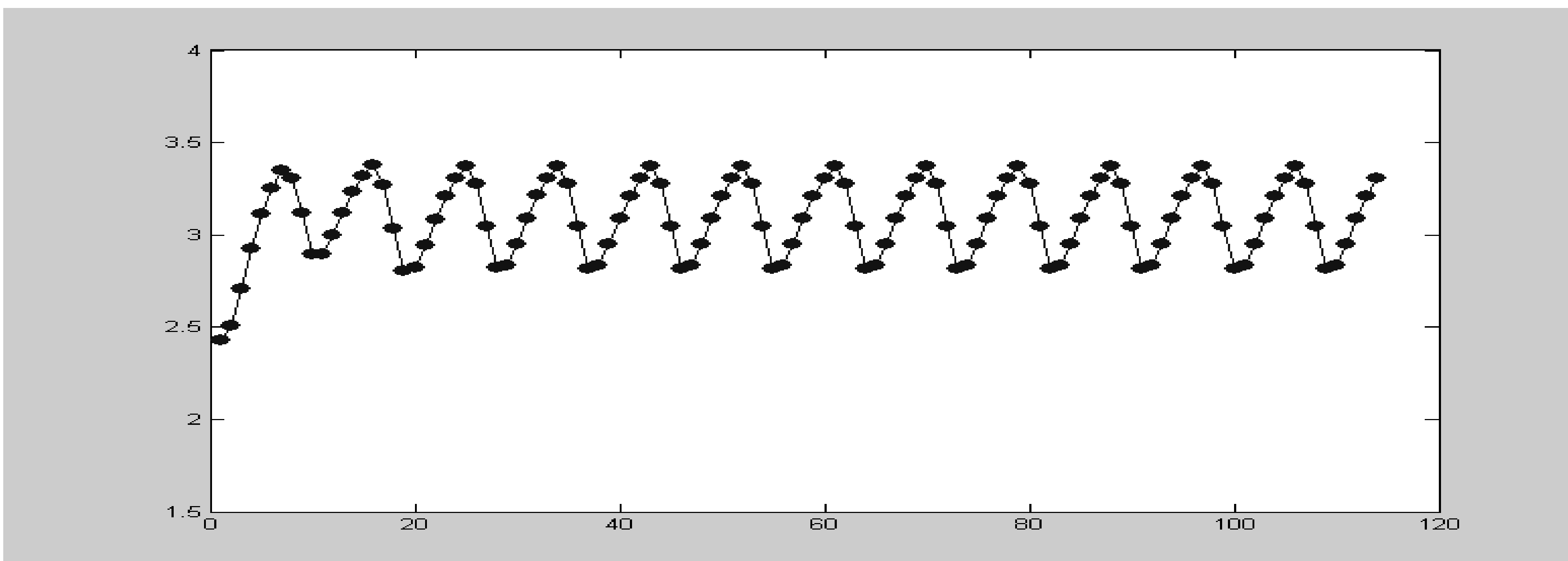
$$x_n = \begin{cases} 0.62 + 1.25 x_{n-1} - 0.43 x_{n-2} + \sigma_1 \varepsilon_n & \text{for } x_{n-2} \le 3.25 \\ 2.25 + 1.52 x_{n-1} - 1.24 x_{n-2} + \sigma_2 \varepsilon_n & \text{for } x_{n-2} > 3.25 \end{cases}$$

- Discarding the innovation term, this model can be rewritten as:

$$x_n - x_{n-1} = \begin{cases} 0.62 + 0.25 x_{n-1} - 0.43 x_{n-2} & \text{for } x_{n-2} \le 3.25 \\ 0.52 x_{n-1} - (1.24 x_{n-2} - 2.25) & \text{for } x_{n-2} > 3.25 \end{cases}$$

- There is a nice ecological interpretation of this model in terms of predator-prey interaction.
- The lower regime corresponds to the increase phase, and the upper one to the decrease phase. Note that the (positive) coefficient of x_{n-1} is smaller in the first regime, while that of x_{n-2} is more negative in the second one.
- This difference in the two phases is called *phase-dependence* in ecology.

- It is also interesting t note that iteration of the deterministic part of the model gives rise to a limit cycle with the same feature of asymmetry.



- In the case that the unobserved threshold variable z_n is a Markov random variable, it is possible to perform a maximum-likelihood estimation, not only on the AR coefficients and variances, but also on the transition and static probabilities of z_n, as well as the
- We will focus on the 2-regime case (2 Markov states) .

- Note that if π_{ij} is the transition probability of z_n from state i to state j, then only π_{11} and π_{22} need be specified, since $\pi_{12} = 1 - \pi_{11}$, and $\pi_{21} = 1 - \pi_{22}$.
- Let Ω_{n-1} be the full information up to time n-1, $\boldsymbol{a}_k = [a_{k0}, \ldots, a_{kp}]^{\mathsf{T}}$ and σ the coefficient vector and (common) standard deviation. Under the assumption the residuals are Gaussian, the density of x_n conditional on z_n and Ω_{n-1} is Gaussian.

- With θ the full parameter vector, and $\boldsymbol{x}_{n\text{-}1} = [1, x_{n\text{-}1}, \ldots, x_{n\text{-}p}]^\mathsf{T}$, this probability density is expressed as:

$$f(x_n \mid z_n = k, \Omega_{n-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_n - \boldsymbol{a}_{\boldsymbol{k}}^\mathsf{T} \boldsymbol{x}_{n-1})^2}{2\sigma^2} \right]$$

- As the state z_n is not observed, the log-likelihood is computed only conditionally on $\Omega_{n\text{-}1}$, i.e. one considers $\ln[f(x_n | \Omega_{n\text{-}1}; \theta)]$.

- The probability density $f(x_n| \Omega_{n-1};\theta)$ can be obtained from:

$$f(x_n | \Omega_{n-1};\theta) = f(x_n, z_n = 1 | \Omega_{n-1};\theta) + f(x_n, z_n = 2 | \Omega_{n-1};\theta)$$

$$= \sum_{k=1}^{2} f(x_n | z_n = k, \Omega_{n-1};\theta) P(z_n = k | \Omega_{n-1};\theta)$$

- The probabilities $P(z_n=k|\Omega_{n-1};\theta)$ are unknown, and must be estimated.

- We are going to need three estimates:

$\hat{\boldsymbol{u}}_{n|n\text{-}1} = [P(z_n\!=\!1|\Omega_{n\text{-}1};\theta)\ P(z_n\!=\!2|\Omega_{n\text{-}1};\theta)]^{\mathsf{T}}$, the *forecast*.

$\hat{\boldsymbol{u}}_{n|n} = [P(z_n\!=\!1|\Omega_n;\theta)\ P(z_n\!=\!2|\Omega_n;\theta)]^{\mathsf{T}}$, the *inference*.

$\hat{\boldsymbol{u}}_{n|t} = [P(z_n\!=\!1|\Omega_t;\theta)\ P(z_n\!=\!2|\Omega_N;\theta)]^{\mathsf{T}}$, the *smoothed inference*.

In the smoothed inference, Ω_N corresponds to all N observations (past and future of n).

- If the state was known at time n-1, then would consist simply of the transition probabilities, that is:

$$\hat{\boldsymbol{u}}_{n|n-1} = \mathbf{P}\boldsymbol{u}_{n-1} = \begin{bmatrix} \pi_{11} & 1-\pi_{22} \\ 1-\pi_{11} & \pi_{22} \end{bmatrix} \boldsymbol{u}_{n-1}$$

with **P** the transition matrix, and $\boldsymbol{u}_{n-1} = [1\ 0]^T$ if $z_{n-1} = 1$ and $\boldsymbol{u}_{n-1} = [0\ 1]^T$ if $z_{n-1} = 2$. Since is not known, it is replaced by an estimate of the probablities of the states at n-1 conditioned on Ω_{n-1}, i.e. $\hat{\boldsymbol{u}}_{n-1|n-1}$

- Given a starting value for $\hat{\boldsymbol{u}}_{1|0}$ (typically two probabilities summing to unity) one can compute the optimal forecast and inference for n = 1,…, N using:

$$\hat{\boldsymbol{u}}_{n|n} = \frac{\hat{\boldsymbol{u}}_{n|n-1} * \boldsymbol{f}_n}{\mathbf{1}^{\mathsf{T}}(\hat{\boldsymbol{u}}_{n|n-1} * \boldsymbol{f}_n)}$$

$$\hat{\boldsymbol{u}}_{n+1|n} = \mathbf{P}\hat{\boldsymbol{u}}_{n|n}$$

Where * denotes element-by-element multiplication , $\mathbf{1} = [1\ 1]^{\mathsf{T}}$ and vector $\boldsymbol{f}_n$ contains the conditional densities of x_n for the two states.

- Now it can be shown that the smoothed inference can be obtained (backwards in time) using:

$$\hat{\boldsymbol{u}}_{n|N} = \hat{\boldsymbol{u}}_{n|n} * (\mathbf{P}^{\mathsf{T}}[\hat{\boldsymbol{u}}_{n+1|N} \div \hat{\boldsymbol{u}}_{n+1|n}])$$

where ÷ denotes element-by-element division.

- The maximum likelihood estimates of the transition probabilities are given by:

$$\hat{\pi}_{ij} = \frac{\sum_{n=2}^{N} P(z_n = j, z_{n-1} = i \,|\, \Omega_n; \hat{\theta})}{\sum_{n=2}^{N} P(z_{n-1} = i \,|\, \Omega_n; \hat{\theta})}$$

- Also, one can obtain the following relationships:

$$\sum_{n=1}^{N} (x_n - \hat{\boldsymbol{a}}_k^{\mathsf{T}} \boldsymbol{x}_{n-1}) x_n P(z_n = k \mid \Omega_n; \hat{\theta}) \qquad k = 1,2$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{2} (x_n - \hat{\boldsymbol{a}}_k^{\mathsf{T}} \boldsymbol{x}_{n-1})^2 P(z_n = k \mid \Omega_n; \hat{\theta})$$

Which mean that the coefficients vectors are obtained as a weighted least square solution, with the weights the square roots of the smoothed probabilities that regime *k* takes place.

- To sum up:

1) starting value of the parameters
2) Computation of the smoothed state probabilities
3) New estimates of transition probabilities
4) New estimates of coefficients vectors and variance
5) Back to 2) until convergence.

Note this algorithm is an expectation-maximization (EM) one, with guaranteed increase of the likelihood.

REFERENCES

1. H. Tong, *Nonlinear Time Series: A Dynamical Systems Approach*, Oxford Univ. Press, Oxford, 1990.
2. P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*, Cambridge Univ. Press, 2000.
3. R.E. McCulloch and R.S. Tsay, “Statistical analysis of economic time series via Markov switching models,” *J. Time Series Anal.*, vol. 15, no. 5, 1994, pp. 523-539.
4. J.D. Hamilton, “Estimation, inference and forecasting of time series subject to changes in regime,” in G.S. Maddala, C.R. Rao, and H.D. Vinod Eds., *Handbook of Statistics,* vol. 11, North-Holland, Amsterdam, pp. 231-260.