Nonparametric Models

Adaptive Functional-Coefficient Autoregressive Models

Fan & Yao, Chapter 8.4

Birgit Schwabl

Outline

- The Functional-Coefficient Autoregressive Model
- **The Adaptive Functional-Coefficient Autoregressive Model**

  - The Model
  - Existence and Identifiability
  - Profile Least-Squares Estimation
  - Bandwidth Selection
  - Variable Selection
  - Implementation

The Functional-Coefficient Autroregressive Model - FAR (p,d)

$$X_t=a_1(X_{t-d})X_1+\ldots+a_p(X_{t-d})X_{t-p}+\sigma(X_{t-d})\varepsilon_t$$

ε_t is a sequence of independent random variables with 0 mean and unity variance.

ε_t is independent of X_{t-1}, X_{t-2}, …

The coefficient functions $a_1(\cdot)$, …, $a_p(\cdot)$ are unknown.

We call X_{t-d} the **model-dependent variable**.

The Functional-Coefficient Autroregressive Model - FAR (p,d)

The FAR model is a natural extension of the TAR model.

It allows the coefficient function to change graduelly, rather than abruptly as in the TAR model, as the value of X_{t-d} varies continuously.

Example:

The population density X_{t-d} changes continuously and its effects on the current population size X_t will be continuous as well.

The Functional-Coefficient Autroregressive Model - FAR (p,d)

The FAR model depends critically on the choice of the model-dependent variable X_{t-d}. This limits the scope of its application.

A generalization of this class of models is to allow a linear combination of past values as a model-dependent variable.

The Functional-Coefficient Autroregressive Model

Let $G(x_1,\ldots, x_p)=E(X_t|X_{t-1}=x_1,\ldots, X_{t-p}=x_p)$
be the autoregressive function. Then, one can write

$$X_t=G(X_{t-1},\ldots, X_{t-p})+\varepsilon_t$$

with $E\{\varepsilon_t|X_{t-1},\ldots, X_{t-p}\}=0$. The autoregressive function G is the best prediction function in the sense that G minimizes the expected prediction error:

$$\min_g E(X_t - g(X_{t-1},\ldots, X_{t-p}))^2$$

The Functional-Coefficient Autroregressive Model

The nonparametric function $G(x_1,\ldots, x_p)$ cannot be estimated with reasonable accuracy due to the curse of dimensionality. Thus, some forms of $G(\cdot)$ are frequently imposed.

The Adaptive Functional-Coefficient Autroregressive Model - AFAR

A generalization of the FAR model is to allow its coefficient functions to depend on the linear combination of past values, called **indices**. Let $\mathbf{X}_{t-1}=(X_{t-1},\ldots, X_{t-p})$ and β be an unknown direction in the p-dimensional space R^p, namely $||\beta||=1$.

The AFAR model approximates the autoregressive function G of the form

$$g(\mathbf{x})=g_0(\beta^T\mathbf{x})+\sum_{j=1}^{p} g_j(\beta^T\mathbf{x})x_j$$

The Adaptive Functional-Coefficient Autroregressive Model - AFAR

When G(**x**) = $g(\mathbf{x}) = g_0(\beta^T x) + \sum_{j=1}^{p} g_j(\beta^T x) x_j$
the AFAR model is given by

$$X_t = g_0(\beta^T \mathbf{X}_{t-1}) + \sum_{j=1}^{p} g_j(\beta^T \mathbf{X}_{t-1}) X_{t-j} + \varepsilon_t$$

ε_t is assumed to be independent of $\mathbf{X}_{t-1}$.

The class of AFAR models is larger than the class of FAR models: Let β be the unit vector with the d*th* position 1 and the rest elements 0, then $\beta^T \mathbf{X}_{t-1} = X_{t-d}$. Thus, the AFAR model includes the FAR model as a specific case.

The AFAR Model: Existence and Identifiability

Questions:

- Does there exist a unique function of the form $g(\mathbf{x})$ such that it minimizes the prediction error $E\{X_t - g(\mathbf{X}_{t-1})\}^2$? ⇨
- Do there exist functions $\{g_j\}$ such that the resulting AFAR model best approximates the autoregressive function G ?

- Is the model $g(\mathbf{x})$ identifiable ?

The AFAR Model: Existence and Identifiability

Answers:

Y is any response variable

$\mathbf{X}=(X_1, \ldots, X_p)^T$ is a vector of predictors

- Assume that ($\mathbf{X}$, Y) has a continuous density and Var(Y)+Var(||X||)<∞. Then, there exists a function g(·) that **minimizes** the prediction error, provided that $\mathrm{Var}(\mathbf{X}^*|\beta^T\mathbf{X})$ is nondegenerate, where $\mathbf{X}^*=(1, X_1, \ldots, X_{p-1})$.

The AFAR Model: Existence and Identifiability

Answers:

- If $=\beta(\beta_1, \ldots, \beta_p)^T$ is given and $\beta_p \neq 0$, then the functions $g_j(\cdot)$ ($j=0, \ldots, p-1$) are uniquely determined from g, i.e. they are **identifiable**.

- For any twice-differentiable $g(\cdot)$, if the first non-zero component of β is chosen to be positive, such a β with $||\beta||=1$ is **unique** unless $g(\cdot)$ is of the form $g(\mathbf{x})=\boldsymbol{\alpha}^T\mathbf{x}\boldsymbol{\beta}^T\mathbf{x}+\boldsymbol{\gamma}^T\mathbf{x}+c$

for some constant vectors α*,* β and γ and a constant c.

The AFAR Model: Profile Least-Squares Estimation

Basic Idea:

First, estimate the nonparametric functions with a given β, resulting in estimates $\widehat{g}_j(\cdot;\beta)$, and then estimate via least-squares method the unknown parameter β using the estimated nonparametric functions $\widehat{g}_j(\cdot;\beta)$.

To ease the computational burden on the nonlinear least-squares estimate to obtain $\widehat{\beta}$ an iterative scheme is frequently employed:

The AFAR Model: Profile Least-Squares Estimation

Basic Idea:

Given an initial estimate $\widehat{\beta}_0$ of β, one obtains estimated coefficient functions $\widehat{g}_j(\cdot;\widehat{\beta}_0)$ and the „synthetic parametric model“

$$g(\mathbf{x}) = \widehat{g}_0(\boldsymbol{\beta}^T\mathbf{x};\widehat{\boldsymbol{\beta}}_0) + \sum_{j=1}^{p-1} g_j(\boldsymbol{\beta}^T\mathbf{x};\widehat{\boldsymbol{\beta}}_0)\,x_j\,.$$

Applying the least-squares method, we obtain a new estimate $\widehat{\beta}_1$. This updates the estimate of β. With updated $\widehat{\beta}_1$, we update the nonparametric components and obtain the estimates $\widehat{g}_j(\cdot;\widehat{\beta}_1)$. With the new nonparametric estimates, we can further update the estimate of parametric component β. Keep iterating until a convergence criterion is met.

The AFAR Model: Bandwidth Selection

The generalized cross-validation (GCV) method is modified to choose the bandwidth h for the estimation of $\{g_j(\cdot)\}$ for a given β.

The optimal bandwidth is $h_{opt}=(a_2/(4na_1))^{1/5}$. The coefficients of a_0, a_1, and a_2 will be estimated from $\{GCV(h_k)\}$ via least-squares regression, where

$$\mathrm{GCV}(h) = a_0 + a_1 h^4 + \frac{a_2}{nh} + o_p(h^4 + n^{-1}h^{-1}).$$

The AFAR Model: Variable Selection

The numbers of predictors in the AFAR model can be large. Hence, it is important to select significant variables.

The basic idea of the local variable selection is as follows:

The least significant variable among X_0, X_1, …, X_{p-1} is deleted from the full model g(**x**), one varible every time, according to its t-value. This yields a sequence of new and reduced models. The best model is selected according to the minimum AIC.

The AFAR Model: Implementation

Step 1: Standardize the data set $\{X_t\}$ such that it has sample mean 0 and the sample variance and covariance matrix I_p. Specify an initial value of β.

Step 2: For a given direction β, we estimate the functions $g_j(\cdot)$. For given $g_j(\cdot)$`s, we update the direction β.

Step 3: Calculate the bandwidth $\widehat{h}$ with β equal to its estimated value.

Step 4: For $h = \widehat{h}$ selected in step 3, repeat step 2 until two successive values of the function to be minimized differ insignificantly.

Step 5: For $\beta = \widehat{\beta}$ obtained from step 4, apply the stepwise deletion technique to select the right variables.

Thank you very much for your attention!