Model and Variable Selection Procedures for Semiparametric Time Series Regression

加藤　リサ (東京理科大学大学院　工学研究科・院)
塩浜　敬之 (東京理科大学大学院　工学研究科)

Semiparametric time series regression models are very useful for econometric and finantial time series analysis. In this paper we proposed the penalized weighted east squares estimators which can simultaneously select the variables and estimation of unknown parameters. The generalized information criterion for model selection is also proposed. We illustrate the effectiveness of proposed procedures with numerical simulations.

We consider the semiparametric regression model

$$y(t) = \alpha(t) + \boldsymbol{\beta}'\boldsymbol{x}(t) + \varepsilon(t), \tag{1}$$

where $y(t)$ is the response variable and $\boldsymbol{x}(t)$ is the $d \times 1$ covariate vector at time t, $\alpha(t)$ is an unspecified baseline function of t, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, and $\varepsilon(t)$ is a Gaussian and zero mean covariance stationary process. Denote by $\gamma(k) = \mathrm{cov}(\varepsilon(t), \varepsilon(t+k)) \equiv \sigma^2\rho_k$. $\alpha(t)$ is expressed as a linear combination of a set of m underlying basis functions of the form

$$\alpha(t, w) = \sum_{k=1}^{m} w_k\phi_k(t) = \boldsymbol{w}'\boldsymbol{\phi}(t),$$

where $\{\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_m(t))'\}$ is un m-dimensional vector constructed from a radial basis functions $\{\phi_k(t); k = 1, \ldots, m\}$, and $\boldsymbol{w} = (w_1, \ldots, w_m)$ is un unknown parameter vector to be estimated. We shall use Gaussian basis function, such that

$$\phi_k(t) = \exp\left(-\frac{\|t - \mu_j\|^2}{2s^2}\right),$$

where μ_j determines the location of the basis function, s^2 determines the width. Then the semiparametric regression model (1) can be expressed as a linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{w} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{y} = (y(1), \ldots, y(n))'$, $\boldsymbol{B} = (\boldsymbol{\phi}_1(1), \ldots, \boldsymbol{\phi}_m(n))'$. Clearly penalized weighted least squares estimator is a minimizer of the following function

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{w})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{w}) + \alpha\boldsymbol{w}'\boldsymbol{K}\boldsymbol{w}, \tag{2}$$

where ξ is the smoothing parameter controlling the trade-off between the goodness-of-fit measure by weighted least squares and the roughness of the estimated function. The most efficient choice of the working matrix is the inverse of the true covariance matrix $\mathbf{\Sigma}$. By simple calculus, (2) is minimized by when $\boldsymbol{\beta}$ and $\boldsymbol{w}$ satisfy the block matrix equation

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{B} \\ \boldsymbol{B}'\boldsymbol{X} & \boldsymbol{B}'\boldsymbol{B}+\alpha\boldsymbol{K} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}' \\ \boldsymbol{B}' \end{pmatrix} \boldsymbol{y}. \tag{3}$$

The equation (3) to be solved without any iteration. First we find $\hat{\boldsymbol{\alpha}} = \boldsymbol{B}\hat{\boldsymbol{w}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, where $\boldsymbol{S} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B} + \alpha\boldsymbol{K})^{-1}\boldsymbol{B}'$ is usually called smoothing matrix. Substituting $\hat{\boldsymbol{\alpha}}$ into (3), we obtain $\hat{\boldsymbol{y}} = \hat{\boldsymbol{X}}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, where $\hat{\boldsymbol{y}} = (\boldsymbol{I}-\boldsymbol{S})\boldsymbol{y}$, $\hat{\boldsymbol{X}} = (\boldsymbol{I}-\boldsymbol{S})\boldsymbol{X}$, and $\boldsymbol{I}$ is the identity matrix of order n. Applying least squares to the linear model, we obtain $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}\hat{\boldsymbol{y}}$. The estimator $\hat{\boldsymbol{\beta}}_n$ is a SOLSE of $\boldsymbol{\beta}$. Speckman (1988) studied similar solutions for partial linear models with independent observations. Since the errors are serially correlated in model (1), the SOLSE is not asymptotically efficient. To obtain an asymptotically efficient estimator for $\boldsymbol{\beta}$, we use the Gram-Schmidt orthgonalization. A nice property about the $\varepsilon^*(t)$ is that they are uncorrelated with constant varianve σ^2. Let $\boldsymbol{T}$ be the $n \times n$ transformation matrix, i.e., $\boldsymbol{\varepsilon}^* = \boldsymbol{T}\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}^* = (\varepsilon^*(1), \ldots, \varepsilon^*(n))$. Then our SGLSE of $\boldsymbol{\beta}$ is obtained by regression $\hat{\boldsymbol{T}}\hat{\boldsymbol{y}}$ on $\hat{\boldsymbol{T}}\hat{\boldsymbol{X}}$, which gives

$$\hat{\boldsymbol{\beta}}_T = (\hat{X}'\hat{T}'\hat{T}\hat{X})^{-1}\hat{X}'\hat{T}'\hat{T}\hat{y} = (\hat{X}\hat{V}^{-1}\hat{X})^{-1}\hat{X}'\hat{V}^{-1}\hat{y}, \tag{4}$$

where $\hat{V}^{-1} = \hat{\sigma}^{-2}\hat{T}'\hat{T}$.

Model selection is an indispensable tool for statistical data analysis. However, it has rarely been studied in the semiparametric context. Fan and Li (2004) studied the penalized weighted least squares estimation with variable selection in semiparametric models for longitudinal data. Their estimation procedure does not include the variable selection for the parameters of baseline functions. The finate sample performances of proposed estimator is also discussed. The penalized weighted least squares estimation procedures works well, that is the proposed procedures effectively reduce model complexity.

Refferences

Fan, J. and Li, R. (2001).*J. Am. Stat. Assoc.*, **96**, 1348-1360.

Fan, J. and Li, R. (2004). *J. Am. Stat. Assoc.*, **99**, 710-723.

Speckman, P. (1988).*J. Royal Statist. Soc. Ser. B*, **50**, 413-436.