2014年6月13日　統計数理研究所　オープンハウス

Penalized Likelihood Estimation in High-Dimensional Time Series Models

植松　良公　統計的機械学習研究センター　日本学術振興会特別研究員（PD）

1 Introduction

Aim: Construct a general estimation method for high-dim. time series models by penalized QML that gives sparse estimates.

Examples: K-dim. VAR(r) model is defined by

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_r y_{t-r} + \varepsilon_t, \tag{1}$$

which has $K^2 r$ parameters. K-dim. MGARCH(1,1) is given by

$$y_t = \Sigma_t^{1/2}\varepsilon_t, \quad \Sigma_t = CC^\top + A^\top y_{t-1}y_{t-1}^\top A + B^\top \Sigma_{t-1} B,$$

which has $K(5K+1)/2$ parameters.

2 General Theory

2.1 The model and its PQML estimator

Model: Let $\{y_t\}_{t=1}^T$ be a vector stationary process with a continuous conditional density $g(y_t|y_{t-1}, y_{t-2}, \ldots)$. Consider a parametric family of densities $\{f(y_t|y_{t-1}, y_{t-2}, \cdots : \theta) : \theta \in \Theta\}$ s.t.:

- $p := \dim(\theta) = O(n^\delta)$ for some $\delta > 0$, so possibly $p > n$;
- the "true value" θ^0, the unique minimizer of the KLIC of g relative to f, is sparse.

Define some notation more precisely:

- $\mathscr{M}_0 = \{j \in \{1, \ldots, p\} : \theta_j^0 \neq 0\}$ and $\mathscr{M}_0^c = \{1, \ldots, p\} \backslash \mathscr{M}_0$;
- $\theta^0_{\mathscr{M}_0}$ is the q-dim. subvector of θ^0 composed of the nonzero elements $\{\theta_j^0 : j \in \mathscr{M}_0\}$;
- $\theta^0_{\mathscr{M}_0^c}$ is the $(p-q)$-dim. subvector of θ^0 composed of zeros.

Estimator: The PQML estimator $\hat\theta$ of θ^0 is defined by

$$Q_n(\hat\theta) = \max_{\theta\in\Theta} Q_n(\theta) \ \text{ with } \ Q_n(\theta) := L_n(\theta) - P_n(\theta),$$

where $L_n(\theta) := n^{-1}\sum_{t=1}^n \log f(y_t|Y_{t-1} : \theta)$ is the quasi-log-likelihood and $P_n(\theta) := \sum_{j=1}^p p_\lambda(|\theta_j|)$ is the penalty term such as L_1-penalty (lasso), SCAD, MCP, etc., with $\lambda(=\lambda_n) \to 0$.

2.2 Theoretical results

Theorem 1 (Weak oracle property) *Under regularity conditions, there is a local maximizer $\hat\theta = (\hat\theta^\top_{\mathscr{M}_0}, \hat\theta^\top_{\mathscr{M}_0^c})^\top$ of $Q_n(\theta)$ s.t.: (a) $P(\hat\theta_{\mathscr{M}_0^c} = 0) \to 1$; (b) $\|\hat\theta_{\mathscr{M}_0} - \theta^0_{\mathscr{M}_0}\|_\infty = O_p(n^{-\gamma}\log n)$.*

Corollary 1 (L_1-penalized QML estimator) *Under regularity conditions in Theorem 1, there is a local maximizer $\hat\theta = (\hat\theta^\top_{\mathscr{M}_0}, \hat\theta^\top_{\mathscr{M}_0^c})^\top$ of $Q_{L_1 n}(\theta)$ s.t. Thm. 1 (a) and (b) hold.*

Theorem 2 (Oracle property) *Under regularity conditions, there is a local maximizer $\hat\theta = (\hat\theta^\top_{\mathscr{M}_0}, \hat\theta^\top_{\mathscr{M}_0^c})^\top$ of $Q_n(\theta)$ s.t.: (a) $P(\hat\theta_{\mathscr{M}_0^c} = 0) \to 1$; (b) $\|\hat\theta_{\mathscr{M}_0} - \theta^0_{\mathscr{M}_0}\| = O_p(n^{-1/2})$. If a stronger assumption is added to the penalty, we have (c) (Asy. N) $n^{1/2}\left(\hat\theta_{\mathscr{M}_0} - \theta^0_{\mathscr{M}_0}\right) \to_d N\left(0,\ (J^0_{\mathscr{M}_0})^{-1} I^0_{\mathscr{M}_0} (J^{0\top}_{\mathscr{M}_0})^{-1}\right)$.*

3 Application to VAR

3.1 Theoretical result for VAR

Consider (1) with $\varepsilon_t \sim$ i.i.d. $(0, \Sigma_\varepsilon)$. Let $\theta^0 = \mathrm{vec}(\Phi_1^0, \ldots, \Phi_r^0) \in \mathbb{R}^p$ with $p = K^2 r$, which is supposed sparse. Using some appropriate Σ instead of unknown Σ_ε, we have:

Proposition 1 *Under some moment and stability conditions, Thm. 2 (a) – (c) hold for $\hat\theta$ in (1), where $I^0_{\mathscr{M}_0} = P^\top_{\mathscr{M}_0}(\Gamma \otimes \Sigma^{-1}\Sigma_\varepsilon\Sigma^{-1})P^\top_{\mathscr{M}_0}$ and $J^0_{\mathscr{M}_0} = P^\top_{\mathscr{M}_0}(\Gamma \otimes \Sigma^{-1})P_{\mathscr{M}_0}$ with $\Gamma = \mathrm{E}[x_t x_t^\top]$.*

3.2 Empirical study

Compare performances of sparse VAR and dynamic Nelson-Siegel (DNS) model in terms of yield curve forecasting.

Data: Zero-coupon US government bond yields that are:

- monthly from January 1986 to December 2007;
- made of 8 maturities $\tau = 3, 6, 12, 24, 36, 60, 84, 120$ months.

Model 1: DNS model is defined by

$$y_{\tau t} = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\eta_t\tau}}{\eta_t\tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\eta_t\tau}}{\eta_t\tau} - e^{-\eta_t\tau}\right),$$
$$\beta_{it} = a_i + b_i\beta_{i,t-h} + u_{it} \quad \text{for each } \ i = 1, 2, 3.$$

where β_{1t}, β_{2t} and β_{3t} may be interpreted as latent dynamic factors and η_t is a sequence of tuning parameters.

Model 2: In sVAR strategy, the model is specified as 8-dim. VAR(12) below and is estimated by SCAD penalized QML.

$$\begin{pmatrix} \Delta y_{3,t} \\ \Delta y_{6,t} \\ \vdots \\ \Delta y_{120,t} \end{pmatrix} = \Phi_1 \begin{pmatrix} \Delta y_{3,t-1} \\ \Delta y_{6,t-1} \\ \vdots \\ \Delta y_{120,t-1} \end{pmatrix} + \cdots + \Phi_{12} \begin{pmatrix} \Delta y_{3,t-12} \\ \Delta y_{6,t-12} \\ \vdots \\ \Delta y_{120,t-12} \end{pmatrix} + \varepsilon_t.$$

Forecasting strategy: The two models are estimated recursively, using the data from Jan. 1986 to the time that the $h(=1, 3, 6, 12)$-month-ahead forecast is made, beginning in Jan. 2001 and extending through Dec. 2007.

Result: The comparison result is summarized below:

Table 1: Relative RMSEs of forecasting (sVAR/DNS)

$h \setminus \tau$	3	6	12	24	36	60	84	120
1	0.356	0.301	0.288	0.279	0.266	0.254	0.258	0.275
3	0.418	0.393	0.358	0.345	0.333	0.324	0.329	0.356
6	0.557	0.513	0.443	0.405	0.391	0.379	0.381	0.400
12	0.625	0.591	0.540	0.492	0.468	0.442	0.435	0.445