Semi-parametrický prístup k odhadovaniu koeficientov ARMA modelov časových radov

Peter Laník

Peter.Lanik@umb.sk
KM FPV, UMB, Tajovského 40, SK 974 01 Banská Bystrica

ABSTRAKT

Pri odhadovaní koeficientov ARMA modelov stacionárnych časových radov sa klasické odhadovacie metódy opierajú o predpokad znalosti pravdepodobnostného rozdelenia šumu, väčšinou gaussovského. V príspevku je prezentovaný prístup, ktorý sa neobmedzuje na jedno konkrétne rozdelenie, ale pripúšťa, že dáta môžu pochádzať aj z iného rozdelenia. Cieľom prístupu je robustnosť voči nesprávnej špecifikácii modelu z hľadiska rozdelenia šumu a zároveň voči aditívnym outlierom.

ÚVOD

Zaoberáme sa slabo stacionárnými, invertibilnými náhodnými procesmi $\{Y_t; t \in Z\}$, $Y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$, kde ϵ_t je biely šum.

Klasické odhadovacie metódy

Metóda maximálnej vierohodnosti (ML) sa pri odhadovaní koeficientov ARMA modelov rutinne používa za predpokladu gaussovského rozdelenia bieleho šumu ϵ_t. Pozorovania do času T sú realizáciou $T-$rozmerného gaussovského rozdelenia. Vierohodnostná funkcia je združená hustota pravdepodobnosti $T-$rozmerného vektora pozorovaní.

Podmienená metóda maximálnej vierohodnosti (CML) predpokladá znalosť počiatočných hodnôt $\{y_0, y_{-1}, \ldots, y_{-p+1}, \epsilon_0, \epsilon_{-1}, \ldots, \epsilon_{-q+1}\}$. Na základe pozorovaní do času T sa problém maximalizácie logaritmu vierohodnostnej funkcie (za predpokladu gaussovského rozdelenia bieleho šumu ϵ_t) redukuje na problém minimalizácie sumy štvorcov $\sum_{t=1}^{T} \epsilon_t^2$.

GMM

Funkcie $u(Y_t, \theta) : V \times \Theta \to R^q$, kde $Y_t \in V$ je stavový priestor procesu $\{Y_t\}$ a $\theta \in \Theta \subseteq R^p$ je priestor parametrov, nazývame odhadovacie funkcie a rovnice $E[u(Y_t), \theta] = 0$ odhadovacie rovnice. Príkladom odhadovacích rovníc môžu byť skórové rovnice ML odhadu. GMM odhadom nazveme tú hodnotu θ, ktorá minimalizuje kvadratickú formu $Q_T(\theta) = T^{-1} \sum_{t=1}^{T} u(Y_t, \theta)^T W_T T^{-1} \sum_{t=1}^{T} u(Y_t, \theta)$, kde W_T je náhodná pozitívne semi-definitná matica váh, ktorá konverguje podľa pravdepodobnosti k pozitívne definitnej matici $q \times q$ konštánt.

SEMI-PARAMETRICKÝ MODEL

Pomocou odhadovacých rovníc vytvoríme teoretický model:

$$\begin{aligned} \Phi(\Theta) &= \bigcup_{\theta \in \Theta} \Phi(\theta), \\ \text{kde } \Phi(\theta) &= \{f_Y(y;\theta) : E[u(Y_t, \theta)] = 0\}. \end{aligned}$$

Na spojenie modelu $\Phi(\Theta)$ s dátami $\{y_t\}_1^T$ vytvoríme empirický model:

$$\begin{aligned} \Phi_T(\Theta) &= \bigcup_{\theta \in \Theta} \Phi_T(\theta), \\ \text{kde } \Phi_T(\theta) &= \{f_Y(y;\theta) : T^{-1} \sum_{t=1}^{T} u(Y_t, \theta) \simeq 0\}. \end{aligned}$$

Z $\Phi_T(\Theta)$ vyberáme $\Phi_T(\theta)$ tak, aby sme minimalizovali $Q_T(\theta)$. Základnou myšlienkou prístupu je vziať za odhadovacie rovnice, skórové rovnice odvodené z CML pre rôzne pravdepodobnostné rozdelenia šumu.

Príklad AR(1)

Majme proces, ktorý sa riadi AR(1) modelom a odhadovacie rovnice odvodíme z skórových rovníc CML pre gaussovské rozdelenie šumu a studentovo s jedným stupňom voľnosti. Odhadovacie funkcie budú mať tvar:

$$u(Y_t;\theta) = \begin{pmatrix} e_t \\ e_t(Y_t - \mu) \\ e_t/(1+e_t^2) \\ e_t(Y_t-\mu)/(1+e_t^2) \end{pmatrix}; e_t = Y_{t+1} - \mu - a_1(Y_t - \mu).$$

Prvé dve sú odvodené z gaussovského CML a posledné dve sú odvodené z studentovho CML s jedným stupňom voľnosti.

Odhad váhovej matice $\widehat{W}_T$

Váhovú maticu na základe informácie do času T odhadneme:

$$\begin{aligned} \widehat{W}_T &= \widehat{\Sigma}_T^{-1} \\ \widehat{\Sigma}_T &= T^{-1} \sum_{t=1}^{T} u(Y_t;\theta) u(Y_t, \theta)^T. \end{aligned}$$

Odhad váhovej matice v našom semi-parametrickom modeli musíme doplniť o podmienku $cov(u_i; u_j) = 0$, **u_i je odvodené z iného CML ako u_j.**

SIMULÁCIE

Scenár simulácií je prevzaný z [4]. Simuluje sa gaussovský časový rad dĺžky 50. Stredná hodnota μ sa odhaduje ako výberový medián, rozptyl šumu σ^2 ako výberový rozptyl reziduálov. Zvyšné koeficienty sa odhadujú pomocou semi-parametrického modelu. Následne sa v časovom rade nahradia prostredné dve pozorovania aditívnymi outliermi a postup sa opakuje. Simulácie majú 10 000 replikácii.

	odľahlé pozorovania					
metóda	bez	[5,5]	[5,0]	[0,5]	[0,0]	[-5,5]
ML	0.0182	0.0115	0.0520	0.0517	0.0196	0.2567
SPM	0.0205	0.0182	0.0505	0.0481	0.0224	0.1065

Tabuľka 1: MSE parametra AR(1) $a_1 = 0.5$

Poďakovanie Rád by som poďakoval MFF UK v Prahe za podporu účasti na Robuste 2012.

Literatúra

[1] Hall, A.R. (2005) *Generalized method of moments*, Oxford University Press, USA.
[2] Hamilton J.D. (1994) *Time series analysis*, Princeton Univ Pr.
[3] Owen A.B. (2001) *Empirical likelihood*, CRC press.
[4] Politis D.N. (2009) *An algorithm for robust fitting of autoregressive models* Economics Letters, 102, 2, 128–131.