AR(1) MODEL BASED ESTIMATION OF DISCRETE HIGH FREQUENCY DATA

Fiori Lamprinakou[1], Petros Dellaportas[1,2] and Omiros Papaspiliopoulos[3]

[1]Department of Statistics, Athens University of Economics and Business, Greece, [2]Department of Statistical Science, University College London, UK, [3] ICREA & Department of Economics and Business Universitat Pompeu Fabra, Spain

Abstract

We deal with discrete time series models appropriate for high frequency intra-day asset returns. We explore Bayesian inference via MCMC and SMC for the observation-driven model of Rydberg and Shephard (1998) and we propose alternative parameter-driven model

The AD model

Let y_i be the price change (in tick) associated with the ith transaction which takes on values -1, 1 or 0, if the price moves one tick down, one tick up or does not move. Following Rydberg and Shephard (1998), y_i can be written as

$$y_i = A_i D_i,$$

where A_i is called activity factor which takes on a value of one, if the price change has taken place and is zero otherwise, D_i is cAed direction factor and takes on values 1,-1 if the price moves up, down respectively.

Obsevation-driven models for the factors

$Z_i|\mathcal{F}_{i-1} \overset{iid}{\sim} \text{Bernoulli}(\pi_i)$, where $\pi_i \triangleq \pi(Z_i = 1|\mathcal{F}_{i-1})$ satisfying:

GLARMA model: $\text{logit}(\pi_i) = \boldsymbol{x}_{i-1}^\top \boldsymbol{\beta} + g_i$, $g_i = \phi g_{i-1} + \delta \varepsilon_{i-1}$, where $|\phi| < 1$, $\delta \in \mathcal{R}$

Autologistic model: $\text{logit}(\pi_i) = \boldsymbol{x}_{i-1}^\top \boldsymbol{\beta} + \zeta_1 Z_{i-2} + \zeta_2 Z_{i-1}$.

We analyse the models with mle, AM (Haario et al. (2001), Roberts and Rosenthal, 2006) and IBIS algorithm (Chopin, 2002).

Notes: Z_i equals A_i or D_i for the activity or direction factor, respectively. $\mathcal{F}_i$ denotes the information set available at the time transaction i takes place

AR(1) model for the factors

$Z_i|\alpha_i, \boldsymbol{x}_{i-1} \overset{iid}{\sim} \text{Bernoulli}(\pi_i)$, where $\pi_i \triangleq \pi(Z_i = 1|\alpha_i, \boldsymbol{x}_{i-1})$ and π_i, α_i satisfying:

Centered model (C): $\text{logit}(\pi_i) = R_i$, $R_i = \boldsymbol{x}_{i-1}^\top \boldsymbol{\beta} + \alpha_i$, $\alpha_i = \phi \alpha_{i-1} + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, $\alpha_1 \sim (0, \sigma^2/(1-\phi^2))$, $|\phi| < 1$.

Non centered model (NC):

$\text{logit}(\pi_i) = \boldsymbol{x}_{i-1}^\top \boldsymbol{\beta} + \sigma \alpha_i$, $\alpha_i = \phi \alpha_{i-1} + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$, $\alpha_1 \sim (0, 1/(1-\phi^2))$, $|\phi| < 1$.

Bayesian analysis: We analyse the components of the state vector one at a time ('O') and in a single move ('A') using the idea of Titsias (2011). The unknown parameters are updated one by one conditional on the current value of the state path. Following Yu & Meng (2011), the models are estimated by interweaving the two strategies at each iteration in order to improve mixing.

Simulation

We simulate 1e+4 data from the C with $\beta_0 = 3$ and ϕ, $\tau = 1/\sigma^2$ vary on $\{0.95, 0.85, -0.65, -0.85\} \times \{0.05, 0.2, 0.5, 1, 2, 10\}$. We use 8e+4 MCMC draws after a burn-in of 2e+4. Priors: $\beta_0 \sim N(0, 10^3)$, $0.5(\phi+1) \sim U(0.5, 0.5)$ and $\tau \sim Gamma(1e{-}3, 1e{+}3)$. Table shows the best model based on the ess per second of the posterior samples.

τ \ ϕ	0.95	0.85	-0.65	-0.85
0.05	CNC(O)	CNC(O)	CNC(O)	CNC(O)
0.2	NCC(O)	CNC(O)	NCC(O)	NCC(O)
0.5	NCC(O)	NCC(O)	CNC(O)	CNC(O)
1	CNC(O)	NCC(A)	NCC(O)	NCC(O)
2	CNC(O)	NCC(A)	NCC(A)	NCC(A)
10	NCC(A)	NCC(A)	NCC(A)	NCC(A)

Real data

- E-mini S&P 500 futures contract, May 16th 2011 to May 24th 2011,
- first five days for estimation, and the rest for prediction
- analyze time periods 9 a.m.-1 p.m. and 1 p.m.-5 p.m. separately.
- sevelar microstructure lag-1 and 2 covariates on the two best observed quote levels

The models have approximately equal predictive log-likelihood. For the direction process, during the morning parameter ϕ is significant while during the afternoon is not. It is better to consider covariates in the activity time than in the trading time.