Quasi-maximum likelihood estimation for multivariate CARMA processes

Eckhard Schlemm

Institute for Advanced Study, Technische Universität München

2^{nd} Northern Triangular Seminar
Stockholm, 17 March 2010

Outline

Introduction

Versatile class of auto-regressive moving-average processes

$$X_n - \varphi_1 X_{n-1} - \ldots - \varphi_p X_{n-p} = \varepsilon_n + \theta_1 \varepsilon_{n-1} + \ldots + \theta_q \varepsilon_{n-q}$$

Extensions to

- multivariate models (Vector ARMA)
- continuous-time models (CARMA)

Advantages:

- Modelling of dependent time series
- High-frequency and/or irregularly spaced observations

Problem: Estimation

Multivariate CARMA processes

$\mathbb{R}^m$-valued Lévy process $\boldsymbol{L}$ satisfying $\mathbb{E}||\boldsymbol{L}(1)||^2 < \infty$.

An $\mathbb{R}^d$-valued second-order MCARMA(p,q) process solves

$$P(D)\boldsymbol{Y}(t) = Q(D)D\boldsymbol{L}(t), \quad D \equiv \frac{d}{dt}.$$

Auto-regressive polynomial

$$P(z) \coloneqq I_d z^p + A_1 z^{p-1} \ldots + A_p \in M_d(\mathbb{R}[z])$$

Moving-average polynomial

$$Q(z) \coloneqq B_0 z^q + B_1 z^{q-1} \ldots + B_q \in M_{d,m}(\mathbb{R}[z])$$

Multivariate CARMA processes

Stationary solution to continuous-time state space model

$$\text{state equation} \quad d\boldsymbol{X}(t) = \mathscr{A}\boldsymbol{X}(t)dt + \mathscr{B}d\boldsymbol{L}(t)$$
$$\text{observation equation} \quad \boldsymbol{Y}(t) = [I_d, 0_d, \ldots, 0_d]\,\boldsymbol{X}(t),$$

$$\mathscr{A} = \begin{bmatrix} 0 & I_d & 0 & \ldots & 0 \\ 0 & 0 & I_d & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & I_d \\ -A_p & -A_{p-1} & \ldots & \ldots & -A_1 \end{bmatrix},$$

$$\mathscr{B} = \begin{bmatrix} \beta_1^T & \cdots & \beta_p^T \end{bmatrix}^T, \quad \beta_{p-j} = -I_{[0:q]}(j)\left[\sum_{i=1}^{p-j-1} A_i\beta_{p-j-i} + B_{q-j}\right]$$

State space models I

General N-dimensional continuous-time state space model:

$$\begin{aligned} \text{state equation} \quad d\boldsymbol{X}(t) =& A\boldsymbol{X}(t)dt + Bd\boldsymbol{L}(t) \\ \text{observation equation} \quad \boldsymbol{Y}(t) =& C\boldsymbol{X}(t), \end{aligned}$$

$$A \in M_N(\mathbb{R}), \quad B \in M_{N,m}(\mathbb{R}), \quad C \in M_{d,N}(\mathbb{R})$$

Sufficient condition for stationarity of the state process $\boldsymbol{X}$:

$$\operatorname{Re}\lambda_\nu < 0, \quad \lambda_\nu,\ \nu = 1, \ldots, N, \text{ eigenvalues of } A.$$

State space models II

$\boldsymbol{X}$ satisfies

$$\boldsymbol{X}(t) = e^{A(t-s)}\boldsymbol{X}(s) + \int_s^t e^{A(t-u)} B d\boldsymbol{L}(u)$$

The output process $\boldsymbol{Y}$ satisfies

$$\boldsymbol{Y}(t) = \int_{-\infty}^t C e^{A(t-u)} B d\boldsymbol{L}(u).$$

Its spectral density is given by

$$f_{\boldsymbol{Y}}(\omega) = \frac{1}{2\pi} C(i\omega - A)^{-1} B \Sigma^{\boldsymbol{L}} B^T (-i\omega - A^T)^{-1} C^T.$$

Equivalence of MCARMA und multivariate state space models

Theorem

The stationary solution $\boldsymbol{Y}$ of the multivariate state space model $(A, B, C, \boldsymbol{L})$ is an $\boldsymbol{L}$-driven MCARMA process with auto-regressive polynomial P and moving-average polynomial Q if and only if

$$C(zI_N - A)^{-1}B = P(z)^{-1}Q(z), \quad \forall z \in \mathbb{C}.$$

A useful decomposition

Theorem

Let $\boldsymbol{Y}$ *be the output process of the SSM* $(A, B, C, \boldsymbol{L})$,

- $A \in M_N(\mathbb{R})$
- $\lambda_i \in \sigma(A)$, $\lambda_i \neq \lambda_j$

$\exists$ *vectors* $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N \in \mathbb{C}^m \backslash \{\boldsymbol{0}_m\}$ *and* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N \in \mathbb{C}^d \backslash \{\boldsymbol{0}_d\}$ *such that*

$$\boldsymbol{Y}(t) = \sum_{\nu=1}^{N} \boldsymbol{Y}_\nu(t), \quad \boldsymbol{Y}_\nu(t) = \boldsymbol{b}_\nu \int_{-\infty}^{t} e^{\lambda_\nu (t-u)} d\langle \boldsymbol{s}_\nu, \boldsymbol{L}(u) \rangle.$$

Probabilistic properties of the sampled process

We observe the process $\boldsymbol{Y}$ at discrete, equally spaced times

$$\boldsymbol{Y}_n^{(h)} := \boldsymbol{Y}(nh), \quad n \in \mathbb{Z}, \quad h > 0.$$

Linear innovations

$$\underset{\text{orthogonal projection onto } \overline{\operatorname{span}}\left\{\boldsymbol{Y}_\nu^{(h)} : -\infty < \nu < n\right\}}{\boldsymbol{\varepsilon}_n^{(h)} = \boldsymbol{Y}_n^{(h)} - \underset{\uparrow}{P_{n-1}} \boldsymbol{Y}_n^{(h)}}, \quad (\boldsymbol{\varepsilon}_n^{(h)})_{n\in\mathbb{Z}} \sim \text{white noise}$$

We define the polynomial

$$\varphi(z) = \prod_{\nu=1}^{N} \left[1 - e^{-\lambda_\nu h} z\right] \in \mathbb{C}[z].$$

VARMA structure of $\boldsymbol{Y}^{(h)}$

Theorem

There exists a stable monic polynomial $\Theta \in M_d(\mathbb{C}[z])$ *of degree at most* $N-1$ *such that*

$$\varphi(B)\boldsymbol{Y}_n^{(h)} = \Theta(B)\boldsymbol{\varepsilon}_n^{(h)}, \qquad B^j\boldsymbol{Y}_n^{(h)} = \boldsymbol{Y}_{n-j}^{(h)},$$

holds.

$\Rightarrow \boldsymbol{Y}^{(h)}$ *is a weak* $\mathrm{VARMA}(N, N-1)$ *process.*

The innovations process

First r eigenvalues real: $\lambda_\nu \in \mathbb{R}$ for $1 \leq \nu \leq r$;
$\lambda_\nu = \overline{\lambda_{\nu+1}} \in \mathbb{C}\backslash\mathbb{R}$ for $\nu = r+1, r+3\ldots, N-1$.

$$M_\nu = \int_0^h e^{(h-u)\lambda_\nu} d\boldsymbol{L}(u)$$

$$\boldsymbol{M} = \left[M_1^T \cdots M_r^T, \mathrm{re}\, M_{r+1}^T, \mathrm{im}\, M_{r+1}^T \cdots \mathrm{re}\, M_{N-1}^T, \mathrm{im}\, M_{N-1}^T\right]^T$$

Theorem

If $\boldsymbol{M}$ has a non-trivial absolutely continuous component with respect to λ^{mN} the innovations process $\varepsilon^{(h)}$ is strongly mixing with exponentially decaying coefficients.

Parameter identifiability in MCARMA models I

Quasi-maximum likelihood approach + discrete observations
$\Rightarrow$ distinction between models based only on second-order properties of the sampled process

Definition (Identifiability)

A collection of continuous-time stochastic processes $(\boldsymbol{Y}_{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta} \in \Theta)$ is **identifiable** if for any $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2$ the two processes $\boldsymbol{Y}_{\boldsymbol{\vartheta}_1}$ and $\boldsymbol{Y}_{\boldsymbol{\vartheta}_2}$ have different spectral densities.
It is h-**identifiable**, $h > 0$, if for any $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2$ the two processes $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_1}$ and $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_2}$ have different spectral densities.

Parameter identifiability in MCARMA models II

$$\psi : \mathbb{R}^q \supset \Theta \ni \boldsymbol{\vartheta} \mapsto (A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}}), \quad \Theta \text{ compact.}$$

Assumption (Minimality)

For all $\boldsymbol{\vartheta} \in \Theta$ the triple $(A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}})$ is minimal in the sense

$$C(zI_m - A)^{-1}B = C_{\boldsymbol{\vartheta}}(zI_N - A_{\boldsymbol{\vartheta}})^{-1}B_{\boldsymbol{\vartheta}} \Rightarrow m \geq N.$$

Assumption (Eigenvalues)

For all $\boldsymbol{\vartheta} \in \Theta$ the spectrum of $A_{\boldsymbol{\vartheta}}$ is contained in the strip

$$\{z \in \mathbb{C} : -\pi/h < \operatorname{Im} z < \pi/h\}.$$

Parameter identifiability in MCARMA models III

Theorem

Parametrization $\psi : \Theta \supset \boldsymbol{\vartheta} \mapsto (A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}})$

- *identifiable*
- *"Minimality"*
- *"Eigenvalues"*

Then the corresponding collection of output processes $\{\boldsymbol{Y}_{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta} \in \Theta\}$ *is* h*-identifiable.*

Gaussian maximum likelihood estimation I

The QML estimator of $\boldsymbol{\vartheta}$ based on $\boldsymbol{y}^L = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L)$ is

$$\boxed{\hat{\boldsymbol{\vartheta}}^L := \operatorname{argmax}_{\boldsymbol{\vartheta}\in\Theta} \mathscr{L}_{\boldsymbol{\vartheta}}\left(\boldsymbol{y}^L\right)}, \quad \text{true parameter } \boldsymbol{\vartheta}_0 \in \operatorname{int}\Theta,$$

where the Gaussian likelihood function is

$$\mathscr{L}_{\boldsymbol{\vartheta}}\left(\boldsymbol{y}^L\right) \sim \left(\prod_{n=1}^{L} \det V_{\boldsymbol{\vartheta},n}\right)^{-1/2} \exp\left\{-\frac{1}{2}\sum_{n=1}^{L} \boldsymbol{e}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta},n}^{-1} \boldsymbol{e}_{\boldsymbol{\vartheta},n}\right\}$$

and

$$\begin{aligned} \boldsymbol{e}_{\boldsymbol{\vartheta},n} =& \boldsymbol{y}_n - P_{n-1} \boldsymbol{Y}_{\boldsymbol{\vartheta},n}^{(h)}\Big|_{\left\{\boldsymbol{Y}_{\boldsymbol{\vartheta},\nu}^{(h)}=\boldsymbol{y}_\nu : 1\leq\nu<n\right\}}, \\ V_{\boldsymbol{\vartheta},n} =& \mathbb{E}\left[\boldsymbol{e}_{\boldsymbol{\vartheta},n}\boldsymbol{e}_{\boldsymbol{\vartheta},n}^T \,\Big|\, \boldsymbol{Y}_{\boldsymbol{\vartheta},\nu}^{(h)} = \boldsymbol{y}_\nu : 1 \leq \nu < n\right]. \end{aligned}$$

Gaussian maximum likelihood estimation II

The sampled process $\boldsymbol{Y}^{(h)}$ satisfies the discrete-time state space model

$$\begin{aligned}\boldsymbol{X}_n^{(h)} &= e^{Ah}\boldsymbol{X}_{n-1}^{(h)} + \boldsymbol{Z}_n,\\ \boldsymbol{Y}_n^{(h)} &= C\boldsymbol{X}_n^{(h)},\end{aligned}$$

where the i.i.d. sequence $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ is given by

$$\boldsymbol{Z}_n = \int_{(n-1)h}^{nh} e^{A(nh-u)}Bd\boldsymbol{L}(u).$$

- Kalman Filter
- Numerical maximization

QML estimation - Consistency

Assumption: h-identifiable parametrization

Theorem (Strong consistency)

For every sampling interval $h > 0$, the QML estimator $\hat{\boldsymbol{\vartheta}}^L$ is strongly consistent, i.e.

$$\hat{\boldsymbol{\vartheta}}^L \to \boldsymbol{\vartheta}_0 \quad \textit{a.s. as } L \to \infty.$$

QML estimation - Normality

Assumption: $\mathbb{E}||\boldsymbol{L}(1)||^{4+\delta} < \infty$ for some $\delta > 0$.

Theorem (Asymptotic normality)

For every sampling interval $h > 0$, the QML estimator $\hat{\boldsymbol{\vartheta}}^L$ is asymptotically normally distributed, i.e.

$$\sqrt{L}\left(\hat{\boldsymbol{\vartheta}}^L - \boldsymbol{\vartheta}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Omega), \quad \Omega = J(\boldsymbol{\vartheta}_0)^{-1} I(\boldsymbol{\vartheta}_0) J(\boldsymbol{\vartheta}_0)^{-1},$$

$$J(\boldsymbol{\vartheta}) = -\lim_{L\to\infty} \frac{2}{L} \frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T} \log \mathscr{L}_{\boldsymbol{\vartheta}}\left(\boldsymbol{y}^L\right),$$
$$I(\boldsymbol{\vartheta}) = \lim_{L\to\infty} \frac{4}{L^2} \operatorname{Var} \frac{\partial}{\partial\boldsymbol{\vartheta}} \log \mathscr{L}_{\boldsymbol{\vartheta}}\left(\boldsymbol{y}^L\right).$$

Echelon state space parametrizations

- Integer $N > 0$ (McMillan degree)
- Non-negative structure indices $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$ satisfying $\sum \nu_i = N$

Canonical parametrization

$$\psi_{\boldsymbol{\nu}} : \mathbb{R}^{q(\boldsymbol{\nu})} \supset \Theta \ni \boldsymbol{\vartheta} \mapsto (A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}}), \quad A_{\boldsymbol{\vartheta}} \in M_N(\mathbb{R})$$

- h-Identifiability
- Every MCARMA process is obtained for some $\boldsymbol{\nu}$.

Examples of canonical parametrizations

$\boldsymbol{\nu} = (1,1)$ (Ornstein-Uhlenbeck type process), 7 parameters:

$$A_{\boldsymbol{\vartheta}} = \begin{bmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_3 & \vartheta_4 \end{bmatrix}, \quad B_{\boldsymbol{\vartheta}} = \begin{bmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_3 & \vartheta_4 \end{bmatrix}, \quad C_{\boldsymbol{\vartheta}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$\boldsymbol{\nu} = (1,2)$, 10 parameters:

$$A_{\boldsymbol{\vartheta}} = \begin{bmatrix} \vartheta_1 & \vartheta_2 & 0 \\ 0 & 0 & 1 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 \end{bmatrix}, \quad B_{\boldsymbol{\vartheta}} = \begin{bmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_4 + \vartheta_5\vartheta_7 \end{bmatrix}, \quad C_{\boldsymbol{\vartheta}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$\boldsymbol{\nu} = (2,1)$, 11 parameters:

$$A_{\boldsymbol{\vartheta}} = \begin{bmatrix} 0 & 1 & 1 \\ \vartheta_1 & \vartheta_2 & \vartheta_3 \\ \vartheta_4 & \vartheta_5 & \vartheta_6 \end{bmatrix}, \quad B_{\boldsymbol{\vartheta}} = \begin{bmatrix} \vartheta_7 & \vartheta_8 \\ \vartheta_1 + \vartheta_2\vartheta_7 & \vartheta_3 + \vartheta_2\vartheta_8 \\ \vartheta_4 + \vartheta_5\vartheta_7 & \vartheta_6 + \vartheta_5\vartheta_8 \end{bmatrix}, \quad C_{\boldsymbol{\vartheta}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Simulation study I

Normal inverse Gaussian Lévy process $\boldsymbol{L}$ with parameters

$$\delta > 0, \kappa > 0, \boldsymbol{\beta} \in \mathbb{R}^d, \Delta \in M_d^+(\mathbb{R})$$

given by normal mean-variance mixture

$$\boldsymbol{L}(1) \overset{\mathcal{D}}{=} \boldsymbol{\mu} + V\Delta\boldsymbol{\beta} + V^{1/2}N,$$

where

$$V \sim \mathrm{IG}(\delta/\kappa, \delta^2), \quad N \sim \mathcal{N}(0, \Delta).$$

- Pure jump
- Skewed
- Semi-heavy tailed

Simulation Study II

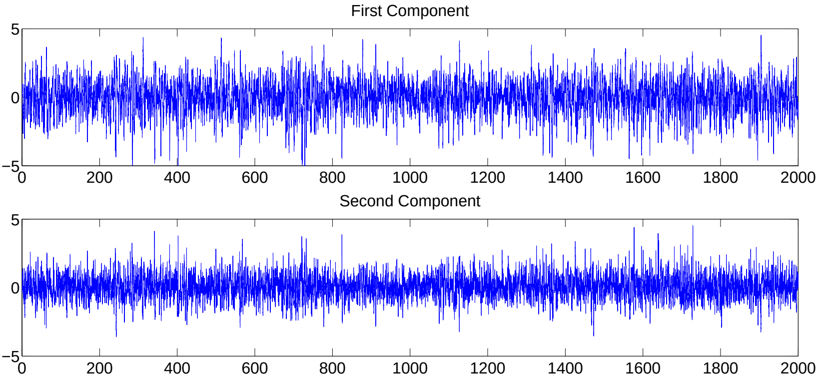
Bivariate NIG-driven $\mathrm{MCARMA}_{1,2}$ process

$$\boldsymbol{X}(t) = \begin{bmatrix} \vartheta_1 & \vartheta_2 & 0 \\ 0 & 0 & 1 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 \end{bmatrix} \boldsymbol{X}(t)dt + \begin{bmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_4 + \vartheta_5\vartheta_7 \end{bmatrix} d\boldsymbol{L}(t),$$

$$\boldsymbol{Y}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{X}(t), \quad \operatorname{vech} \Sigma^{\boldsymbol{L}} = (\vartheta_8, \vartheta_9, \vartheta_{10}).$$

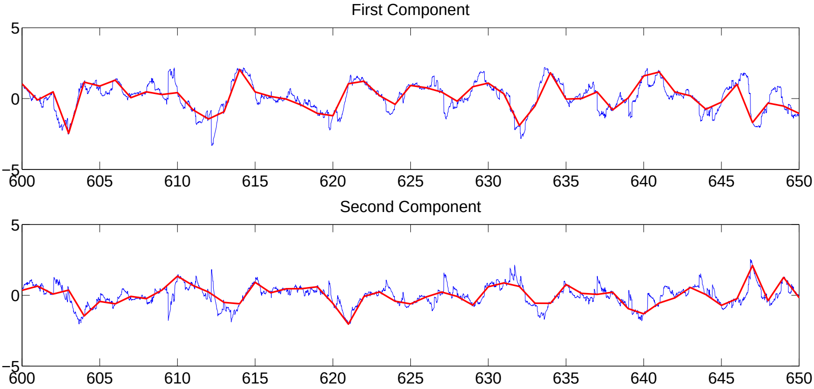
Simulation Study III

One realization of a bivariate NIG-driven $\mathrm{MCARMA}_{1,2}$ process



Simulation Study III

The effect of sampling

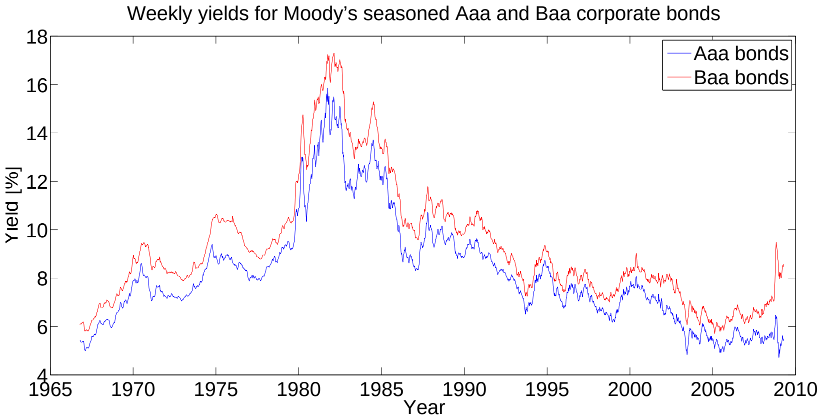


QML estimates for a bivariate NIG-driven $\mathrm{MCARMA}_{1,2}$

- Time horizon $[0, 2000]$
- Observed at integer times
- 350 replicates

para.	sample mean	bias	sample std. dev.	est. std. dev.
ϑ_1	-1.0001	0.0001	0.0354	0.0381
ϑ_2	-2.0078	0.0078	0.0479	0.0539
ϑ_3	1.0051	-0.0051	0.1276	0.1321
ϑ_4	-2.0068	0.0068	0.1009	0.1202
ϑ_5	-2.9988	-0.0012	0.1587	0.1820
ϑ_6	1.0255	-0.0255	0.1285	0.1382
ϑ_7	2.0023	-0.0023	0.0987	0.1061
ϑ_8	0.4723	-0.0028	0.0457	0.0517
ϑ_9	-0.1654	0.0032	0.0306	0.0346
ϑ_{10}	0.3732	0.0024	0.0286	0.0378

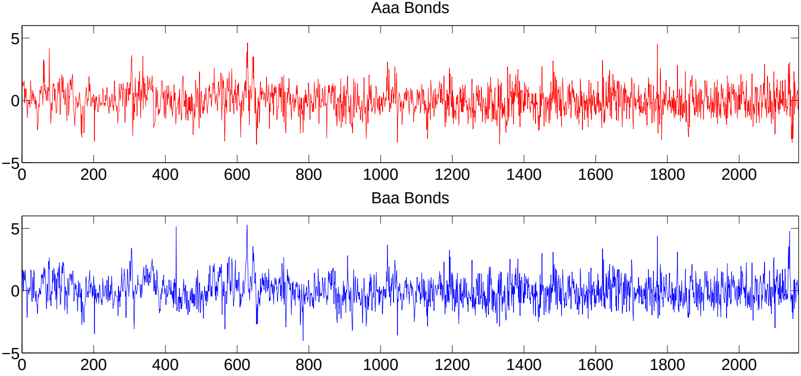
Application to corporate bond yields



Application to corporate bond yields

Data show unit roots but no cointegration

Weekly log-yields after differencing and devolatilization using a moving window of width 52 (corresponding to one year)



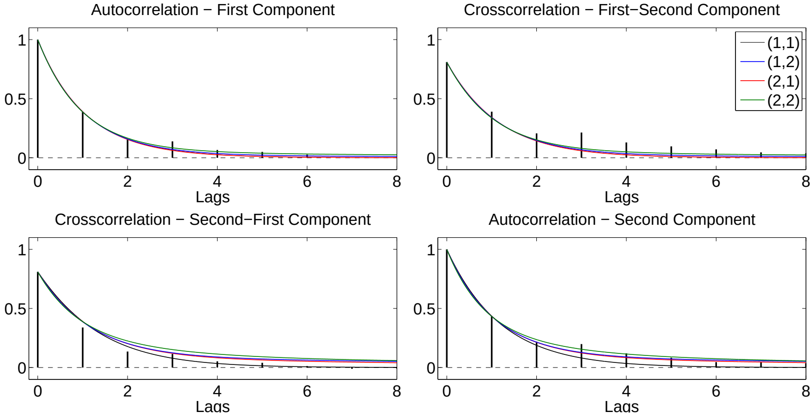
Application to corporate bond yields

QMLE estimates of the parameters of an $MCARMA_{\alpha,\beta}$ model for weekly yields of Moody's seasoned corporate bonds

(α,β)	$(1,1)$		$(1,2)$		$(2,1)$		$(2,2)$	
	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$
$\hat{\vartheta}_1$	-1.1326	0.1349	-1.1538	0.1401	-1.3776	0.0320	**-0.0010**	**0.0336**
$\hat{\vartheta}_2$	0.2054	0.1171	0.2307	0.1008	-2.4033	0.0197	-1.1601	0.5964
$\hat{\vartheta}_3$	0.3316	0.1206	-0.2528	0.1716	0.0228	0.0050	**-0.0098**	**0.0268**
$\hat{\vartheta}_4$	-1.0935	0.1065	**-0.0362**	**0.0472**	-4.9948	0.1096	**0.1829**	**0.7429**
$\hat{\vartheta}_5$	2.4105	0.2324	-1.2516	0.1286	-4.6276	0.1538	1.4646	0.3931
$\hat{\vartheta}_6$	2.2483	0.2061	-2.5747	0.4595	-0.0153	0.0108	1.3662	0.4039
$\hat{\vartheta}_7$	2.7055	0.2116	1.6345	0.2940	-1.2442	0.0391	-0.7438	0.2387
$\hat{\vartheta}_8$			2.8552	0.1966	0.2573	0.0492	-1.7563	0.7209
$\hat{\vartheta}_9$			3.5702	0.2151	2.4302	0.1370	-2.6936	0.6694
$\hat{\vartheta}_{10}$			4.9076	0.3888	2.9784	0.2766	1.7369	0.5381
$\hat{\vartheta}_{11}$					4.1571	0.5043	-3.6136	3.0265
$\hat{\vartheta}_{12}$							2.8483	2.5122
$\hat{\vartheta}_{13}$							4.4848	0.3327
$\hat{\vartheta}_{14}$							5.5079	0.1803
$\hat{\vartheta}_{15}$							7.0218	1.4357
$-2\log L_{\boldsymbol{\vartheta}}(\boldsymbol{y})$	**9,893.8**		**9,850.4**		**9,853.0**		**9,840.7**	

Application to corporate bond yields

Empirical autocorrelations compared to those of the estimated models



Summary and future work

- Quasi-maximum likelihood estimation for second-order MCARMA processes based on discrete observations
- Canonical parametrizations
- Application to a bivariate economic time series

- Preliminary estimates for the numerical likelihood maximization
- Estimation of the driving Lévy process
- Model selection criteria
- Structure of an irregularly sampled MCARMA process