

# Efficient Iterative ML Estimation of High-Parameterized Time Series Models

Nikolaus Hautsch

Ostap Okhrin

Alexander Ristig

Department of Statistics and Operations  
Research

University of Vienna

Ladislaus von Bortkiewicz Chair of Statistics

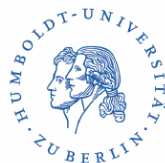
C.A.S.E. – Center for Applied Statistics  
and Economics

Humboldt–Universität zu Berlin

<http://isor.univie.ac.at>

<http://lvb.wiwi.hu-berlin.de>

<http://case.hu-berlin.de>



# VAR

Application: Impulse response analysis.

## Example 1

Let  $x_i$  denote a  $(d \times 1)$  vector of time series variables,  $i = 1, \dots, n$ .

$$x_i = \underbrace{\omega}_{(d \times 1)} + \underbrace{A}_{(d \times d)} x_{i-1} + \varepsilon_i,$$

is known as VAR(1). Least squares estimation is based on moment conditions  $E(\varepsilon_i) = 0$  and  $E(\varepsilon_i \varepsilon_i^\top) = \Sigma_\varepsilon$ .



# VARMA

Application: Forecasting of macroeconomic variables.

## Example 2

Let  $x_i$  denote a  $(d \times 1)$  vector of time series variables,  $i = 1, \dots, n$ .

$$x_i = \underbrace{\omega}_{(d \times 1)} + \underbrace{A}_{(d \times d)} x_{i-1} + \varepsilon_i + \underbrace{B}_{(d \times d)} \varepsilon_{i-1},$$

is known as VARMA(1, 1). Maximum likelihood estimation needs a distribution assumption like  $\varepsilon_i \sim F(0, \Sigma_\varepsilon)$ , with  $E(\varepsilon_i \varepsilon_i^\top) = \Sigma_\varepsilon$ .



## VMEM

Applications: Forecasting of liquidity measures, risk management, volatility contagion.

### Example 3

Let  $x_i \geq 0$  denote a  $(d \times 1)$  vector of time series variables,  $i = 1, \dots, n$ .

$$x_i = \mu_i \odot \varepsilon_i,$$
$$\mu_i = \underbrace{\omega}_{(d \times 1)} + \underbrace{A}_{(d \times d)} x_{i-1} + \underbrace{B}_{(d \times d)} \mu_{i-1},$$

is known as VMEM(1, 1), where “ $\odot$ ” is the component-wise Hadamard product, and  $\varepsilon_{ij} \geq 0$ ,  $j = 1, \dots, d$ . GMM estimation is based on  $\varepsilon_i \sim (1_d, \Sigma_\varepsilon)$  with  $E(\varepsilon_i \varepsilon_i^\top) = \Sigma_\varepsilon$ .



## Copula-GARCH

Applications: VaR-estimation, asset pricing.

### Example 4

Let  $x_i$  denote a  $(d \times 1)$  time series variables,  $i = 1, \dots, n$ .

$$x_i = \Sigma_i \varepsilon_i,$$

$$\Sigma_i = \text{diag}(\sigma_{i1}, \dots, \sigma_{id})$$

$$\sigma_{ij}^2 = \omega_j + a_j x_{i-1,j}^2 + b_j \sigma_{i-1,j}^2,$$

is known as copula-GARCH(1, 1), where

$\varepsilon_i \sim F_{\varepsilon_i}(\varepsilon_1, \dots, \varepsilon_d) = C\{F_{\varepsilon_{i1}}(\varepsilon_1), \dots, F_{\varepsilon_{id}}(\varepsilon_d)\}$  with  $E(\varepsilon_i) = 0$ .



## Related to practitioners

- Volatility contagion via connectedness measures
- Asset and option pricing
- Estimation of VaR and ES
- Forecasting of macroeconomic variables
- Modeling of liquidity measures
- ...



## Challenges for large $d$

- *Non*-Gaussian white noise with a *non*-elliptical dependence structure.
  - ▶ High-dimensional copulae, see Smith et al. (2010, JASA) and Okhrin et al. (2013, JoE).
- Complexity of log-likelihood
  - ▶ Iterative maximization of parts of the log-likelihood, see Song et al. (2005, JASA).
  - ▶ Decomposition of the parameter space in order to update the estimator.
  - ▶ Analytical first-order derivatives of the entire log-likelihood are not required.



# Outline

1. Motivation ✓
2. Efficient estimation
3. Simulation I
4. Sparse and efficient estimation
5. Simulation II
6. Iterative Generalized Least Squares Estimation
7. Application
8. Summary



## An iterative estimation procedure

- Let  $X = (X_1^\top, \dots, X_n^\top)^\top$  be the finite history of the  $d$ -dimensional stochastic process  $\{X_i\}_{i=1,2,\dots}$ .
- Each  $X_i$  has conditional density  $f_{X_i|\mathcal{F}_{i-1}}(\cdot; \vartheta)$ .
- W.l.o.g. decompose  $\vartheta = \mathbf{v}(\vartheta_1, \dots, \vartheta_G) \stackrel{\text{def}}{=} (\vartheta_1^\top, \dots, \vartheta_G^\top)^\top$ , s.t.

$$\begin{aligned} \ell_i(\vartheta) &= \log f_{X_i|\mathcal{F}_{i-1}}(X_{i1}, \dots, X_{id}; \vartheta) \\ &= \sum_{j=1}^d \log f_{X_{ij}|\mathcal{F}_{i-1}}(X_{ij}; \vartheta_1, \dots, \vartheta_k) \\ &\quad + \log c_{X_i|\mathcal{F}_{i-1}} \{ F_{X_{i1}|\mathcal{F}_{i-1}}(X_{i1}; \vartheta_1, \dots, \vartheta_k), \\ &\quad \dots, F_{X_{id}|\mathcal{F}_{i-1}}(X_{id}; \vartheta_1, \dots, \vartheta_k); \vartheta_{k+1}, \dots, \vartheta_G \}. \end{aligned}$$



- Construct the log-likelihood

$$\begin{aligned}\mathcal{L}(\vartheta) &= \sum_{i=1}^n \ell_i(\vartheta) \\ &= \sum_{i=1}^n \{\ell_i^m(\vartheta_1, \dots, \vartheta_k) + \ell_i^c(\vartheta)\} \\ &= \mathcal{L}^m(\vartheta_1, \dots, \vartheta_k) + \mathcal{L}^c(\vartheta).\end{aligned}$$

- Shorthand notation, e.g.,

$$\dot{\mathcal{L}}(\vartheta_0) = \left. \frac{\partial \mathcal{L}(\vartheta)}{\partial \vartheta} \right|_{\vartheta=\vartheta_0}.$$



## Algorithm

*Step 1:*

$$(1) \quad (\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1) = \arg \underset{\vartheta_1, \dots, \vartheta_k}{\text{zero}} \dot{\mathcal{L}}^m(\vartheta_1, \dots, \vartheta_k)$$

$$(2) \quad (\vartheta_{k+1,n}^1, \dots, \vartheta_{G,n}^1) = \arg \underset{\vartheta_{k+1}, \dots, \vartheta_G}{\text{zero}} \dot{\mathcal{L}}^c(\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1, \vartheta_{k+1}, \dots, \vartheta_G)$$



## Algorithm

Step 1:

$$(1) \quad (\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1) = \arg \underset{\vartheta_1, \dots, \vartheta_k}{\text{zero}} \dot{\mathcal{L}}^m(\vartheta_1, \dots, \vartheta_k)$$

$$(2) \quad (\vartheta_{k+1,n}^1, \dots, \vartheta_{G,n}^1) = \arg \underset{\vartheta_{k+1}, \dots, \vartheta_G}{\text{zero}} \dot{\mathcal{L}}^c(\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1, \vartheta_{k+1}, \dots, \vartheta_G)$$

Step  $h > 1$ :

$$(1) \quad \vartheta_{1,n}^h = \arg \max_{\vartheta_1} \mathcal{L}(\vartheta_1, \vartheta_{2,n}^{h-1}, \dots, \vartheta_{G,n}^{h-1})$$

$$(2) \quad \vartheta_{2,n}^h = \arg \max_{\vartheta_2} \mathcal{L}(\vartheta_{1,n}^h, \vartheta_2, \vartheta_{3,n}^{h-1}, \dots, \vartheta_{G,n}^{h-1})$$

⋮

$$(G) \quad \vartheta_{G,n}^h = \arg \max_{\vartheta_G} \mathcal{L}(\vartheta_{1,n}^h, \dots, \vartheta_{G-1,n}^h, \vartheta_G)$$



## Asymptotic properties

### Theorem

Let the random vectors of the sequence  $X$  have an identical conditional density  $f_{X_i|\mathcal{F}_{i-1}}(\cdot; \vartheta)$  for which Assumptions 1-2 hold. If  $\vartheta_n^1 \xrightarrow{P} \vartheta_0$ , then  $\vartheta_n^h \xrightarrow{P} \vartheta_0, \forall h = 2, 3, \dots$

► Assumptions



## Theorem

Let the random vectors of the sequence  $X$  have an identical conditional density  $f_{X_i|\mathcal{F}_{i-1}}(\cdot; \vartheta)$  for which Assumptions 1-4 hold. Then,

$$n^{1/2}(\vartheta_n^h - \vartheta_0) \xrightarrow{\mathcal{L}} \mathbf{N}\{0, \mathcal{B}_h \Sigma(\vartheta_0) \mathcal{B}_h^\top\}.$$

▶ Assumptions

and

▶ Definitions

## Corollary

Under Assumptions 1-4,

$$\lim_{h \rightarrow \infty} n^{1/2}(\vartheta_n^h - \vartheta_0) \xrightarrow{\mathcal{L}} \mathbf{N}\{0, \mathcal{J}(\vartheta_0)^{-1}\}.$$



## Properties under misspecification

Under certain regularity assumptions, see White (1994):

- $\vartheta_n^h$  is a consistent estimator for  $\vartheta_n^*$  - the minimizer of the Kullback-Leibler divergence.
- $n^{1/2}(\vartheta_n^h - \vartheta_n^*)$  converges to a multivariate Gaussian distribution as  $n \rightarrow \infty$ .
- The asymptotic covariance of  $\lim_{h \rightarrow \infty} n^{1/2}\vartheta_n^h$  collapses to

$$\{\mathcal{H}(\vartheta_n^*)^{-1}\} \mathcal{T}_2 \Sigma(\vartheta_n^*) \mathcal{T}_2^\top \{\mathcal{H}(\vartheta_n^*)^{-1}\}^\top.$$



## Setup I

Similar to Kascha (2012, Econometric Reviews):

$$x_i = Ax_{i-1} + \varepsilon_i + B\varepsilon_{i-1}.$$

- $d = 5$ ,  $n = 50$ ,  $r = 24$
- Replication: 500
- $\varepsilon_{ij}$  follow  $t_{\nu_j}$  margins coupled with a correlation matrix  $\vartheta_G$  of a Gaussian copula with  $G = 4$ .
- Decomposition

$$\vartheta_1 = (\nu_1, \dots, \nu_d)^\top,$$

$$\vartheta_2 = \text{vec}(A),$$

$$\vartheta_3 = \text{vec}(B).$$





## Results

Figure 1: The updated mean of the centered estimates of degrees of freedom  $\nu_{j,n}^h - \nu_j$  (solid line) and  $\nu_{j,n}^{h-1} - \nu_j$  (dashed line),  $j = 1, \dots, d$ .

Efficient Iterative ML Estimation



Figure 2: The updated mean squared error for  $A_n^h$  and  $B_n^h$ .



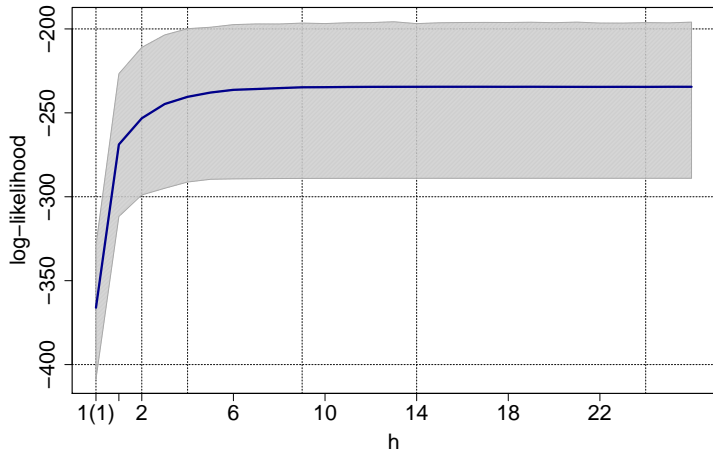


Figure 3: The **median** of the log-likelihood for each step of the iteration. The gray area contains 95% of the sample.



## Penalized 2-stage ML estimation

- Curse of dimensionality
  - ▶ Need to balance the trade-off between few parameters per  $\vartheta_g$ ,  $g = 1, \dots, G$ , and a large  $G$ !
  - ▶ Parameter shrinkage via nonconcave penalized likelihood, see Fan and Li (2001, JASA).
  
- First derivative of the SCAD penalty

$$p'_{\lambda,a}(x) = \lambda \mathbf{I}(x \leq \lambda) + \max(a\lambda - x, 0) / (a - 1) \mathbf{I}(x > \lambda),$$

$a > 2$  and  $x > 0$ . [▶ Simulation](#)



- Split the parameters into
  - ▶ penalized parameters  $\vartheta_{p_m} \stackrel{\text{def}}{=} \mathbf{v}(\vartheta_1, \vartheta_2)$  and  $\vartheta_{p_c} \stackrel{\text{def}}{=} \mathbf{v}(\vartheta_{G-1}, \vartheta_G)$  and
  - ▶ non-penalized parameters  $\vartheta_m \stackrel{\text{def}}{=} \mathbf{v}(\vartheta_3, \dots, \vartheta_k)$  and  $\vartheta_c \stackrel{\text{def}}{=} \mathbf{v}(\vartheta_{k+1}, \dots, \vartheta_{G-2})$ .
- Introduce
  - ▶ meaningful penalization targets  $\check{\vartheta}_1, \check{\vartheta}_2, \check{\vartheta}_{G-1}, \check{\vartheta}_G$  and
  - ▶ the modified SCAD-penalty  $\check{p}_{\lambda,a}(\gamma) = p_{\lambda,a}(|\gamma - \check{\gamma}|)$ .
- W.l.o.g. let  $\vartheta_{1,0} = \check{\vartheta}_1$  and  $\vartheta_{G,0} = \check{\vartheta}_G$ , so that  $f_i(\cdot; \vartheta_0)$  has a less complicated functional form than  $f_i(\cdot; \vartheta)$  for  $\vartheta \neq \vartheta_0$ .



The penalized log-likelihoods are

$$\mathcal{L}^{p_m}(\vartheta_1, \dots, \vartheta_k) = \mathcal{L}^m(\vartheta_1, \dots, \vartheta_k) - n \sum_{l=1}^{r_1+r_2} \check{p}_{\lambda_n^m, a^m}(\vartheta_{l, p_m}),$$

$$\mathcal{L}^{p_c}(\vartheta) = \mathcal{L}^c(\vartheta) - n \sum_{l=1}^{r_{G-1}+r_G} \check{p}_{\lambda_n^c, a^c}(\vartheta_{l, p_c}).$$

## Algorithm

Step 1:

$$(1) \quad (\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1) = \arg \underset{\vartheta_1, \dots, \vartheta_k}{\text{zero}} \dot{\mathcal{L}}^{p_m}(\vartheta_1, \dots, \vartheta_k)$$

$$(2) \quad (\vartheta_{k+1,n}^1, \dots, \vartheta_{G,n}^1) = \arg \underset{\vartheta_{k+1}, \dots, \vartheta_G}{\text{zero}} \dot{\mathcal{L}}^{p_c}(\vartheta_{1,n}^1, \dots, \vartheta_{k,n}^1, \vartheta_{k+1}, \dots, \vartheta_G)$$



## Theorem

Let the random vectors of the sequence  $X$  have an identical conditional density  $f_i(\cdot; \vartheta)$  for which Assumptions 1-3 and 5 hold and let the penalty fulfill certain regularity conditions. If  $\lambda_n^m, \lambda_n^c \rightarrow 0$ ,  $n^{1/2}\lambda_n^m \rightarrow \infty$  and  $n^{1/2}\lambda_n^c \rightarrow \infty$  as  $n \rightarrow \infty$ , then,

- (a)  $\vartheta_{1,n}^1 \xrightarrow{a.s.} \check{\vartheta}_1$  and  $\vartheta_{G,n}^1 \xrightarrow{a.s.} \check{\vartheta}_G$ ,
- (b)  $\vartheta_{2,n}^1 + \mathcal{O}(a_n^m) \xrightarrow{P} \vartheta_{2,0}$  and  $\vartheta_{G-1,n}^1 + \mathcal{O}(a_n^c) \xrightarrow{P} \vartheta_{G-1,0}$ , with  $a_n^m, a_n^c \rightarrow 0$  for  $\lambda_n^m, \lambda_n^c \rightarrow 0$  as  $n \rightarrow \infty$ ,
- (c)  $\vartheta_{m,n}^1 \xrightarrow{P} \vartheta_{m,0}$  and  $\vartheta_{c,n}^1 \xrightarrow{P} \vartheta_{c,0}$ .

► Assumptions



## Iterative Efficient and Sparse Estimation

Step  $h > 1$ :

(1) {blank step}

$$(2) \vartheta_{2,n}^h = \arg \max_{\vartheta_2} \mathcal{L}(\check{\vartheta}_1, \vartheta_2, \vartheta_{3,n}^{h-1}, \dots, \vartheta_{G-1,n}^{h-1}, \check{\vartheta}_G)$$

⋮

$$(G-1) \vartheta_{G-1,n}^h = \arg \max_{\vartheta_{G-1}} \mathcal{L}(\check{\vartheta}_1, \vartheta_{2,n}^h, \dots, \vartheta_{G-2,n}^h, \vartheta_{G-1}, \check{\vartheta}_G)$$

(G) {blank step}





## Corollary

*Under the assumptions of Theorem 3. If  $\lambda_n^m, \lambda_n^c \rightarrow 0$ ,  $n^{1/2}\lambda_n^m \rightarrow \infty$  and  $n^{1/2}\lambda_n^c \rightarrow \infty$  as  $n \rightarrow \infty$ ,  $\tilde{\vartheta}_n^h \xrightarrow{P} \tilde{\vartheta}_0 \forall h = 2, 3, \dots$ , where  $\tilde{\vartheta} = \mathbf{v}(\vartheta_2, \dots, \vartheta_{G-1})$ .*

## Corollary

*Under the assumptions of Theorem 2 and Theorem 3. If  $\lambda_n^m, \lambda_n^c \rightarrow 0$ ,  $n^{1/2}\lambda_n^m \rightarrow \infty$  and  $n^{1/2}\lambda_n^c \rightarrow \infty$  as  $n \rightarrow \infty$ , then*

$$n^{1/2}\mathcal{B}_{h,n}^{-1} \left\{ (\tilde{\vartheta}_n^h - \tilde{\vartheta}_0) + \tilde{\Gamma}^{h-1} \mathcal{K}_n \mathbf{b}_n \right\} \xrightarrow{\mathcal{L}} \mathbf{N} \left\{ 0, \Sigma(\tilde{\vartheta}_0) \right\}.$$

► Definitions



## Setup II

$$x_i = \mu_i \odot \varepsilon_i,$$
$$\mu_i = \omega + A x_{i-1} + \text{diag}(b_{11}, \dots, b_{dd}) \mu_{i-1},$$

- $d = 15, n = 500, r = 375.$
- Replications: 500.
- Penalized parameters: 210 off-diagonal elements of  $A$ .
- $\varepsilon_{ij} \sim \text{Weibull}(\gamma_j)$  are contemporaneous dependent via R-vine, see Kurowicka and Joe (2011).
- Decomposition  $\mathbf{v}(\gamma_j, \omega_j, A_{j\bullet}, b_{jj})$  for  $j = 1, \dots, d$ . ▶ Application



# Results

Figure 4: Comparison of the true matrix  $A$  (left) with **one** updated estimate  $A_n^h$  (right).

Efficient Iterative ML Estimation



Figure 5: The updated mean of the centered estimated parameters of the Weibull distributions  $\gamma_{j,n}^h - \gamma_j, j = 1, \dots, d$  (solid line) and the sd illustrated as grey area.



## Numerical criteria

Define for the parameter vector  $z$  and its estimate  $z_n^h$

1. the relative absolute error:

$$\text{RAE}^h \stackrel{\text{def}}{=} \frac{\|z - z_n^h\|_1}{\|z - z_n^1\|_1}$$

2. the sign consistency:

$$\text{SC}^h \stackrel{\text{def}}{=} \sum_{k \neq \ell} \mathbf{I} \left\{ \text{sign}(A_{k\ell,0}) = \text{sign}(A_{k\ell,n}^h) \right\}.$$



$h$	Parameter	RAE <sup><math>h</math></sup>		SC <sup><math>h</math></sup>	
1(1)	$A_{kl}, k \neq l$	0.35	(0.09)	169	(10.38)
2	$A_{kl}, k \neq l$	0.34	(0.10)	169	(10.38)
	$\omega_j, A_{jj}, B_{jj} \forall j$	0.88	(0.17)		-
4	$A_{kl}, k \neq l$	0.32	(0.10)	169	(11.86)
	$\omega_j, A_{jj}, B_{jj} \forall j$	0.82	(0.18)		-
11	$A_{kl}, k \neq l$	0.31	(0.09)	169	(10.38)
	$\omega_j, A_{jj}, B_{jj} \forall j$	0.80	(0.18)		-

Table 1: Median values of RAE <sup>$h$</sup>  and SC <sup>$h$</sup>  for different parameters. The MAD is given in parentheses.



## Selection of $\lambda_n^m$ and $a^m$

- Split  $\{x_i\}_{i=1}^n$  in two parts:  $S_1 = \{x_i\}_{i=1}^{n_1}$  and  $S_2 = \{x_j\}_{j=n_1+1}^n$  containing 80% and 20% of the sample, respectively.
- Use  $S_1$  to estimate  $\vartheta_{1,n}(\lambda, a), \vartheta_{2,n}(\lambda, a)$ , defined through Step 1(1) of Algorithm 2.
- Fit the tuning parameters through

$$(\lambda_n^m, a^m)^\top = \arg \max_{(\lambda, a)^\top} \mathcal{L}^m \{ \vartheta_{1,n}(\lambda, a), \vartheta_{2,n}(\lambda, a), \vartheta_{3,n}, \dots, \vartheta_{k,n} \}$$

on  $S_2$ , where  $\vartheta_{3,n}, \dots, \vartheta_{k,n}$  are the non-penalized estimators..



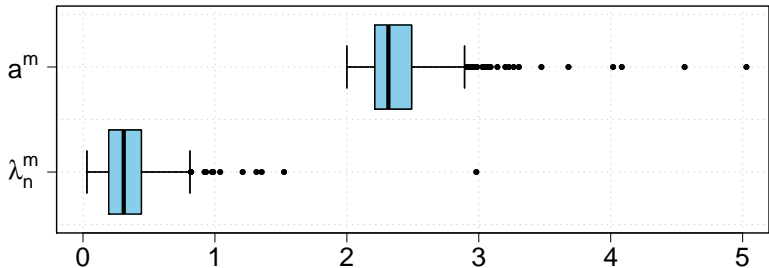


Figure 6: Boxplots for the tuning parameters of the penalization function.

► SCAD penalty





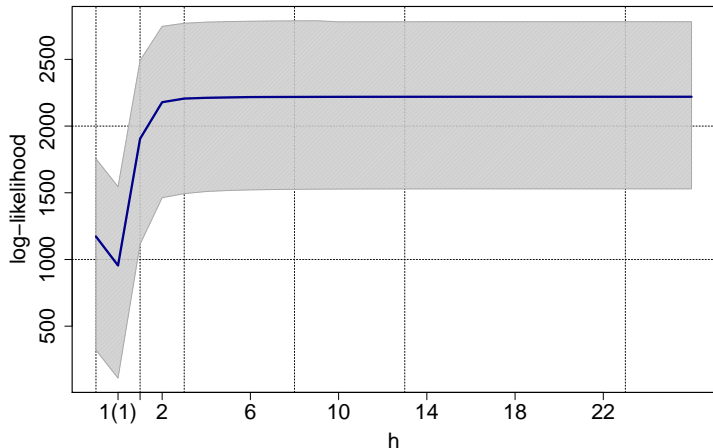


Figure 7: The **median** of the log likelihood for each step of the iteration. The gray area includes 0.95% of the observations.



## VAR

Consider the time series model

$$x_i = c + \sum_{l=1}^q A_l x_{i-l} + \varepsilon_i,$$

where  $c = (c_1, \dots, c_d)^\top$  and  $A_l$  is a  $(d \times d)$  matrix. Given standard assumptions like

- $E(\varepsilon_i \varepsilon_i^\top) = \Sigma_\varepsilon$  and  $E(\varepsilon_i \varepsilon_{i-l}^\top) = 0_{dd}$  for  $l > 0$
- $\varepsilon = \text{vec}(\varepsilon_1, \dots, \varepsilon_d) \sim N(0, I_n \otimes \Sigma_\varepsilon)$

the parameters can be efficiently estimated by OLS. But

- $r > n$  especially for a large  $q$ !



Define  $Y = \text{vec}(x_1, \dots, x_n)$ ,  $Z_i = (1, x_{i-1}^\top, \dots, x_{i-q}^\top)^\top$  and  $Z = (Z_1, \dots, Z_n)$  and rewrite the model in matrix notation

$$Y = (Z^\top \otimes I_d)\beta + \varepsilon,$$

where  $\beta = \text{vec}(c, A_1, \dots, A_q)$ . We assume  $\varepsilon \sim N(0, \Sigma)$ , with  $\Sigma \neq I_n \otimes \Sigma_\varepsilon$ , but the GLS estimator

$$\beta_n = \left\{ (Z \otimes I_d)\Sigma^{-1}(Z^\top \otimes I_d) \right\}^{-1} (Z \otimes I_d)\Sigma^{-1} Y$$

is not feasible.



## Algorithm

*Step 1:*

$$(1) \beta_n^1 = \{(Z Z^T)^{-1} Z \otimes I_d\} Y$$

$$(2) \Sigma_n^1 = \{Y - (Z^T \otimes I_d)\beta_n^1\} \{Y - (Z^T \otimes I_d)\beta_n^1\}^T$$

*Step  $h > 1$ :*

$$(1) \beta_n^h = \{(Z \otimes I_d)(\Sigma_n^{h-1})^{-1}(Z^T \otimes I_d)\}^{-1} (Z \otimes I_d)(\Sigma_n^{h-1})^{-1} Y$$

$$(2) \Sigma_n^h = \{Y - (Z^T \otimes I_d)\beta_n^h\} \{Y - (Z^T \otimes I_d)\beta_n^h\}^T$$

Penalization of  $\beta$  can be embedded at *Step 1!*



## Measuring volatility connectedness

- Daily realized volatilities (RVs) from January 2007 - December 2008.
- 30 U.S. blue chip companies similar to the DJIA.
- VMEM(1, 1) as in Simulation II [▶ VMEM](#).
- R-vine based on bivariate  $t$ -copulae.
- $r/n \approx 1.7$



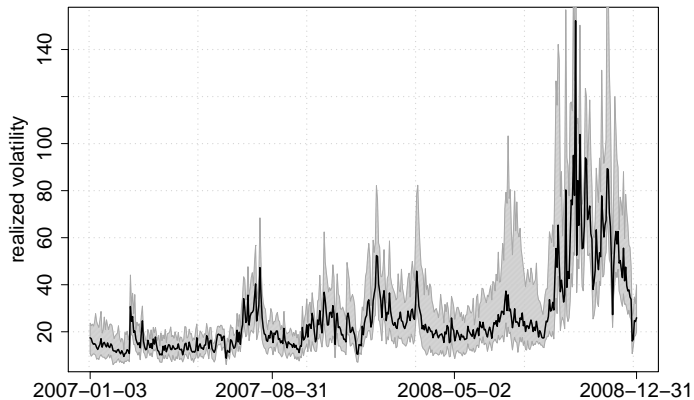


Figure 8: Median of the realized volatilities over the companies. The gray area includes 90% of the observations.



Assuming a stationary VMEM(1, 1) for the RVs  $\{x_i\}_{i=1}^n$ , whose zero-mean MA( $\infty$ ) representation is

$$\begin{aligned}y_i &= \eta_i + \sum_{l=1}^{\infty} \left\{ (A+B)^l - (A+B)^{l-1}B \right\} \eta_{i-l} \\ &= \eta_i + \sum_{l=1}^{\infty} \Psi_l \eta_{i-l},\end{aligned}$$

with  $E(\eta_i) = 0$ ,  $E(\eta_i \eta_i^T) = \Sigma_\eta$  and  $y_i = x_i - \{I_d - (A+B)\}^{-1} \omega$ .

Two types of  $H$ -step prediction errors:

- $\nu_i(H) = \sum_{l=0}^{H-1} \Psi_l \eta_{i+H-l}$  and
- $\nu_{i,\ell}(H) = \sum_{l=0}^{H-1} \Psi_l \left\{ \eta_{i+H-l} - E(\eta_{i+H-l} | \eta_{\ell, i+H-l} = \delta) \right\}$ .



## Connectedness measures

The elements of the generalized variance decomposition matrix  $\tilde{V}_H$  are

$$\tilde{v}_{kl,H} = \frac{e_k^\top [\text{Var} \{v_i(H)\} - \text{Var} \{v_{i,l}(H)\}] e_k}{e_k^\top \text{Var} \{v_i(H)\} e_k},$$

where  $e_k = (0, \dots, 0_{k-1}, 1_k, 0_{k+1}, \dots, 0)^\top$  is a  $(d \times 1)$  vector. Standardization  $v_{kl,H} = \tilde{v}_{kl,H} / \sum_{\ell=1}^d \tilde{v}_{kl,H}$  leads to, see Diebold and Yilmaz (2014, JoE):

- the total directional connectedness to others from  $\ell$  by

$$C_{\bullet \leftarrow \ell, H} = \sum_{k \neq \ell} v_{\bullet k, H},$$

- the total connectedness  $C_H = \sum_{k \neq \ell} v_{kl, H}$ .





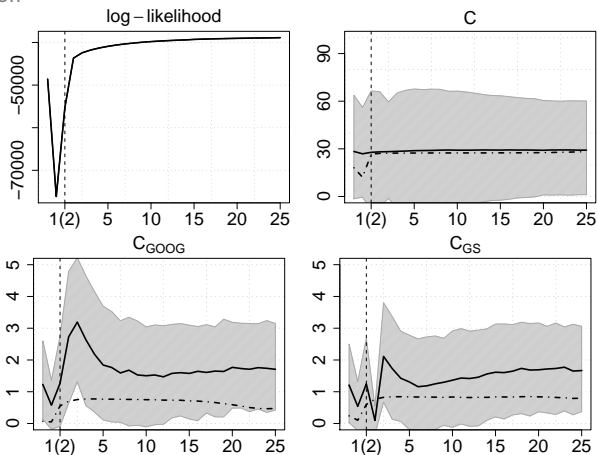


Figure 9: Upper panel: log-likelihood values and total systemic connect-edness  $C_{12}$  in dependence of  $h$ . Lower panel: volatility contagion from Google  $C_{\bullet \leftarrow GOOG,12}$  and Goldman Sachs  $C_{\bullet \leftarrow GS,12}$  in dependence of  $h$ .



## Conclusion

- Maximization strategy for complicated and high-parameterized log-likelihood functions.
- Asymptotic properties of the sparse and efficient estimator are established.
- Accuracy of the procedure is illustrated in a simulation study.
- Application emphasizes the importance of efficiency.

### Future research:

- Hidden Markov models
- Non-parametric margins and semi-parametric efficiency



# Efficient Iterative ML Estimation of High-Parameterized Time Series Models

Nikolaus Hautsch

Ostap Okhrin

Alexander Ristig

Department of Statistics and Operations  
Research

University of Vienna

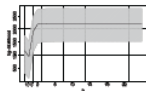
Ladislaus von Bortkiewicz Chair of Statistics  
C.A.S.E. – Center for Applied Statistics  
and Economics

Humboldt–Universität zu Berlin

<http://isor.univie.ac.at>

<http://lvb.wiwi.hu-berlin.de>

<http://case.hu-berlin.de>



## References



Diebold, F. X. and K. Yilmaz

*On the Network Topology of Variance Decompositions:  
Measuring the Connectedness of Financial Firms*  
Journal of Econometrics, forthcoming, 2014



Fan, J. and R. Li

*Variable Selection via Nonconcave Penalized Likelihood and Its  
Oracle Properties*  
Journal of the American Statistical Association, Vol. 96, No.  
456, 1348–1360, 2001



Kascha, C.

*A Comparison of Estimation Methods for Vector Autoregressive  
Moving-Average Models*  
Econometric Reviews 31(3), 297–324, 2012





Kurowicka, D. and H. Joe

*Dependence Modeling: Vine Copula Handbook*

World Scientific Publishing Company, Incorporated, 2011



Okhrin, O., Okhrin, Y. and W. Schmid

*Determining the Structure and Estimation of Hierarchical Archimedean Copulas*

Journal of Econometrics 173(2), 189–204, 2013



Smith, M., Min, A., Almeida, C. and C. Czado

*Modelling Longitudinal Data Using a Pair-copula*

*Decomposition of Serial Dependence*

Journal of the American Statistical Association. 105(492),

1467-1479, 2010





Song, P., Fan, Y. and J. D. Kalbfleisch

*Maximization by Parts in Likelihood Inference*

Journal of the American Statistical Association, Vol. 100, No. 472, 1145–1158, 2005



White, H.

*Estimation, Inference and Specification Analysis*

Cambridge University Press (Cambridge), 1st Edition, 1994



## Assumptions

- (1) The model is identifiable and the true value  $\vartheta_0$  is an interior point of the compact parameter space  $\Theta$ . We assume that the model is correctly specified in the sense that  $E_{\vartheta}\{\partial\ell_i(\vartheta)/\partial\vartheta_g\} = 0$  and information equality holds,

$$\mathcal{I}_{i,gl}(\vartheta) \stackrel{\text{def}}{=} E_{\vartheta} \left\{ \frac{\partial\ell_i(\vartheta)}{\partial\vartheta_g} \frac{\partial\ell_i(\vartheta)}{\partial\vartheta_l^\top} \right\} = - E_{\vartheta} \left\{ \frac{\partial^2\ell_i(\vartheta)}{\partial\vartheta_g\partial\vartheta_l^\top} \right\},$$

for  $g, l = 1, \dots, G$  and  $i = 1, \dots, n$ .



(2) The information matrix is  $\mathcal{I}(\vartheta) = \sum_{i=1}^n \mathcal{I}_i(\vartheta)$ , with  $\mathcal{I}_i(\vartheta) = \{\mathcal{I}_{i,gl}(\vartheta)\}_{g,l=1}^G$ . Let the limit of  $n^{-1}\mathcal{I}(\vartheta) \xrightarrow{P} \mathcal{J}(\vartheta)$  be the asymptotic information matrix, which is finite and positive definite at  $\vartheta_0$  and  $n^{-1}\ddot{\mathcal{L}}(\vartheta) \xrightarrow{P} \mathcal{H}(\vartheta)$  be the asymptotic Hessian, which is finite and negative definite for  $\vartheta \in \{\vartheta : \|\vartheta - \vartheta_0\| < \delta\}$ ,  $\delta > 0$ . [▶ Back](#)

(3) The score  $s(\vartheta_0) = \mathbf{v}\{\dot{\mathcal{L}}^m(\vartheta_{1,0}, \dots, \vartheta_{k,0}), \dot{\mathcal{L}}^c(\vartheta_0)\}$  of the decomposed log-likelihood  $\mathcal{L}(\vartheta) = \mathcal{L}^m(\vartheta_1, \dots, \vartheta_k) + \mathcal{L}^c(\vartheta)$ , with  $\{n^{-1}s(\vartheta_0)s(\vartheta_0)^\top\} \xrightarrow{P} \Sigma(\vartheta_0)$ , obeys

$$n^{-1/2}s(\vartheta_0) \xrightarrow{\mathcal{L}} N\{0, \Sigma(\vartheta_0)\}.$$

[▶ Back](#)



- (4) Define the lower block and upper block triangular matrix of  $-n^{-1}\ddot{\mathcal{L}}(\vartheta_0)$  as  $L_n$  and  $U_n$ , respectively, such that  $-n^{-1}\ddot{\mathcal{L}}(\vartheta_0) = L_n - U_n$  with  $L_{gl,n} = 0$  for  $g < l \leq G$  and  $U_{gl,n} = 0$  for  $l \leq g \leq G$ . For the probability limits  $L$  and  $U$  of  $L_n$  and  $U_n$ , respectively, we assume  $\rho(\Gamma) < 1$ , where  $\rho(\cdot)$  denotes the spectral radius and  $\Gamma \stackrel{\text{def}}{=} L^{-1}U$ . [▶ Back](#)
- (5) There exists an open subset  $\theta$  of  $\Theta$  that contains the true parameter  $\vartheta_0$  such that for almost all  $X_i$ ,  $i = 1, \dots, n$ , the density  $f_i(\cdot; \vartheta)$  admits all third derivatives  $\partial f_i(X_{i1}, \dots, X_{id}; \vartheta) / \partial \vartheta_u \partial \vartheta_v \partial \vartheta_w$  for all  $\vartheta \in \theta$ . Furthermore, there exist functions  $M_{uvw}(\cdot)$  such that

$$\left| \frac{\partial \ell_i(\vartheta)}{\partial \vartheta_u \partial \vartheta_v \partial \vartheta_w} \right| \leq M_{uvw}(X_i) \quad \text{for all } \vartheta \in \theta,$$

where  $E\{M_{uvw}(X_i)\} < \infty$  for  $u, v, w = 1, \dots, r$ . [▶ Back](#)



## Definitions

- Let the number of parameters of each subvector  $\vartheta_g$  be  $r_g$ ,  $g = 1 \dots, G$ , s.t.  $p = \sum_{g=1}^k r_g$  and  $r = \sum_{g=1}^G r_g$ . Define  $q = r - p$  and the matrices

$$\mathcal{T}_1 = \begin{pmatrix} I_p & 0_{pp} & 0_{pq} \\ 0_{qp} & 0_{qp} & I_q \end{pmatrix} \quad \text{and} \quad \mathcal{T}_2 = \begin{pmatrix} I_p & I_p & 0_{pq} \\ 0_{qp} & 0_{qp} & I_q \end{pmatrix},$$

with identity matrix  $I_p$ ,  $0_{pq} = 0_p 0_q^\top$  and null vector  $0_p$ .



□ Define

$$n^{-1} \begin{Bmatrix} \ddot{\mathcal{L}}^m(\vartheta_{1,0}, \dots, \vartheta_{k,0}) 0_{pq} \\ \ddot{\mathcal{L}}^c_{\vartheta(\vartheta_{k+1}, \dots, \vartheta_G), \vartheta}(\vartheta_0) \end{Bmatrix} = \mathcal{H}^1(\vartheta_0) + \mathcal{O}_p(1),$$

and

$$\mathcal{B}_h = \Gamma^{h-1} [\mathcal{K}\mathcal{T}_1 - \{-\mathcal{H}(\vartheta_0)\}^{-1}\mathcal{T}_2] + \{-\mathcal{H}(\vartheta_0)\}^{-1}\mathcal{T}_2,$$

$$\text{and } \mathcal{K} = \{-\mathcal{H}^1(\vartheta_0)\}^{-1}.$$

▶ Back



Define

$$\mathbf{b}_n^m \stackrel{\text{def}}{=} \left\{ \check{\rho}'_{\lambda_n^m, a^m}(\vartheta_{21,0}), \dots, \check{\rho}'_{\lambda_n^m, a^m}(\vartheta_{2r_2,0}) \right\}^\top,$$

$$\mathbf{b}_n^c \stackrel{\text{def}}{=} \left\{ \check{\rho}'_{\lambda_n^c, a^c}(\vartheta_{(G-1)1,0}), \dots, \check{\rho}'_{\lambda_n^c, a^c}(\vartheta_{(G-1)r_{G-1},0}) \right\}^\top.$$

and  $\mathbf{b}_n = (\mathbf{b}_n^m, \mathbf{0}_s, \mathbf{b}_n^c)^\top$  as well as

$$\Psi_n^m = \text{diag} \left\{ \check{\rho}''_{\lambda_n^m, a^m}(\vartheta_{21,0}), \dots, \check{\rho}''_{\lambda_n^m, a^m}(\vartheta_{2r_2,0}) \right\},$$

$$\Psi_n^c = \text{diag} \left[ \check{\rho}''_{\lambda_n^c, a^c} \{ \vartheta_{(G-1)1,0} \}, \dots, \check{\rho}''_{\lambda_n^c, a^c} \{ \vartheta_{(G-1)r_{G-1},0} \} \right].$$

and  $\Psi_n = \text{diag} (\Psi_n^m, \mathbf{0}_{ss}, \Psi_n^c)$ .



The “bread” of the covariance matrix is given by:

$$\mathcal{B}_{h,n} = \tilde{\Gamma}^{h-1} \left[ \mathcal{K}_n \mathcal{T}_1 - \{-\mathcal{H}(\tilde{\vartheta}_0)\}^{-1} \mathcal{T}_2 \right] + \{-\mathcal{H}(\tilde{\vartheta}_0)\}^{-1} \mathcal{T}_2,$$

and  $\mathcal{K}_n = \left\{ \Psi_n - \mathcal{H}^1(\tilde{\vartheta}_0) \right\}^{-1}.$

▶ Back

