Sparse estimation of Vector Autoregressive Models

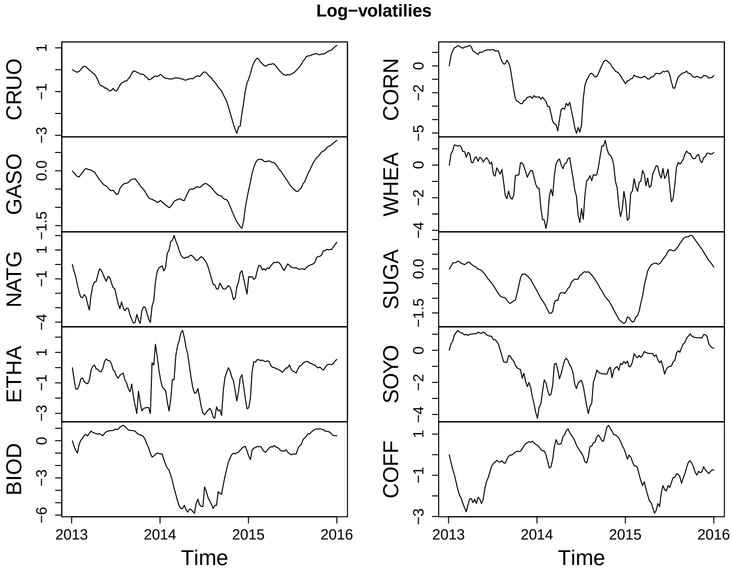
Christophe Croux

KU Leuven

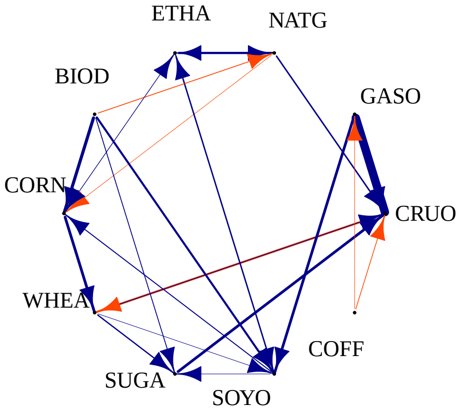
Limassol, 9 April 2017

(joint work with Ines Wilms, Luca Barbaglia, Sarah Gelper)

10-dimensional time series



Network



Vector Autoregressive Model (VAR)

Two stationary time series $y_{1,t}$ and $y_{2,t}$.

VAR(1) in dimension $q = 2$:

$$\begin{cases} y_{1,t} & = & \Gamma_{1,11}\, y_{1,t-1} + \Gamma_{1,12} y_{2,t-1} + e_{1t} \\ y_{2,t} & = & \Gamma_{1,21}\, y_{1,t-1} + \Gamma_{1,22} y_{2,t-1} + e_{2t} \end{cases}$$

Covariance matrix of $(e_{1t}, e_{2t})'$ is Σ.
Vector notation: $\mathbf{y}_t = \Gamma_1 \mathbf{y}_{t-1} + \mathbf{e}_t$,

The VAR model

Let $\mathbf{y}_t$ be a q-dimensional stationary time series

Vector Autoregressive Model of order p:

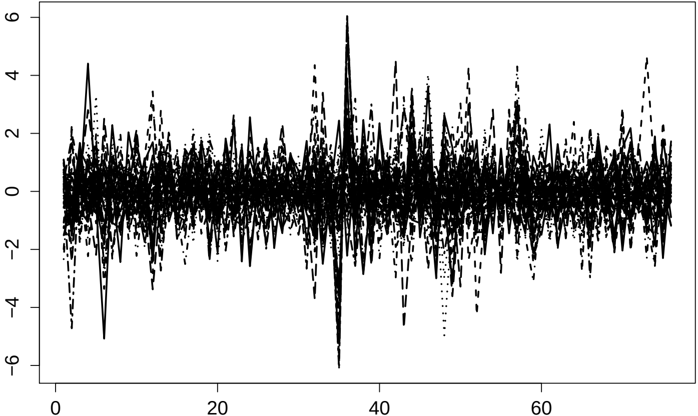
$$\mathbf{y}_t = \Gamma_1 \mathbf{y}_{t-1} + \Gamma_2 \mathbf{y}_{t-2} + \ldots + \Gamma_p \mathbf{y}_{t-p} + \mathbf{e}_t \,,$$

- Matrices Γ_j are autoregressive parameters
- $\mathbf{e}_t$ error with covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$.
- Standard estimation procedure: OLS equation by equation.

Example: a Market Response Model

Sales, promotion and prices for 17 product categories: $q = 51$

$T = 77$ weekly observations



VAR model for $q = 3 \times 17 = 51$ time series

- One lag
  - $1 \times (q \times q) = 2601$ regression parameters
  - 1326 unique elements in $\boldsymbol{\Sigma}$
- Two lags
  - $2 \times (q \times q) = 5202$ regression parameters
  - 1326 unique elements in $\boldsymbol{\Sigma}$

$\rightarrow$ Explosion of number of parameters

The VAR model: Overparametrization

ML estimators will be

- Not computable
- Inaccurate

Sparse estimation $\equiv$ many estimated parameters equal to zero

- Suitable if T is small relative to the number of parameters
- Easier to interpret
- Automatic variable selection
- Better estimation and prediction performance

Sparse Estimation: Lasso

In the multiple linear regression model

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

Minimization problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\,(y - X\beta)'(y - X\beta) + \lambda \sum_{l=1}^{k} |\beta_l|.$$

Tibshirani (1996)

Lasso for the VAR model

- Multiple equations
  - Partial correlation between the error terms
    $\rightarrow$ Glasso of Friedman et al. (2008)
- Dynamic nature of the model
  - Selecting a time series into one of the equations = selecting the variable and all its lags
    $\rightarrow$ Group lasso (Yuan and Lin, 2006)

Penalized ML estimation

Rewrite the VAR in matrix notation:

$$\mathbf{Y} = \mathbf{Y_L}\mathbf{\Gamma} + \mathbf{E},$$

where

- $\mathbf{Y} = (\mathbf{y}_{p+1}, \ldots, \mathbf{y}_T)'$
- $\mathbf{Y_L} = (\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T)'$ with $\mathbf{X}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$
- $\mathbf{\Gamma} = (\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p)'$
- $\mathbf{E} = (\boldsymbol{e}_{p+1}, \ldots, \boldsymbol{e}_T)'$.

Penalized ML estimation (cont.)

Penalized negative log likelihood:

$$(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\Omega}}) = \underset{\boldsymbol{\Gamma}, \boldsymbol{\Omega}}{\text{argmin}} \; \frac{1}{T}\text{tr}\Big((\mathbf{Y} - \mathbf{Y_L}\boldsymbol{\Gamma})\boldsymbol{\Omega}(\mathbf{Y} - \mathbf{Y_L}\boldsymbol{\Gamma})'\Big) - \log|\boldsymbol{\Omega}| + \lambda_1 \sum_{g=1}^{G} ||\gamma_g||_2 + \lambda_2 \sum_{k \neq k'} |\Omega_{kk'}|,$$

with

- γ_g a subvector of $\boldsymbol{\Gamma}$
- $G = q^2$ total number of groups.
- $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ the precision matrix

Algorithm

Solving for $\mathbf{\Gamma}|\mathbf{\Omega}$*:*

$$\widehat{\mathbf{\Gamma}}|\mathbf{\Omega} = \underset{\mathbf{\Gamma}}{\text{argmin}} \ \frac{1}{T}\text{tr}\Big((\mathbf{Y} - \mathbf{Y_L}\mathbf{\Gamma})\mathbf{\Omega}(\mathbf{Y} - \mathbf{Y_L}\mathbf{\Gamma})'\Big) + \lambda_1 \sum_{g=1}^{G} ||\gamma_g||_2.$$

$\rightarrow$ groupwise lasso

Algorithm (cont.)

Solving for $\boldsymbol{\Omega}|\boldsymbol{\Gamma}$*:*

$$\widehat{\boldsymbol{\Omega}}|\boldsymbol{\Gamma} = \underset{\boldsymbol{\Omega}}{\text{argmin}} \;\; \frac{1}{T}\text{tr}\Big((\mathbf{Y}-\mathbf{Y_L}\boldsymbol{\Gamma})\boldsymbol{\Omega}(\mathbf{Y}-\mathbf{Y_L}\boldsymbol{\Gamma})'\Big) - \log|\boldsymbol{\Omega}| + \lambda_2 \sum_{k\neq k'} |\Omega_{kk'}|.$$

$\rightarrow$ penalized inverse covariance estimation (glasso)

Selection of tuning parameters

In the iteration step $\mathbf{\Gamma}|\mathbf{\Omega}$, select λ_1 to minimize

$$BIC_{\lambda_1} = -2 \log L_{\lambda_1} + k_{\lambda_1} \log(T),$$

- L_{λ_1} is the estimated likelihood using λ_1
- k_{λ_1} is the number of non-zero estimated regression coefficients.

In the iteration step $\mathbf{\Omega}|\mathbf{\Gamma}$, select λ_2 analogously.

Networks from the VAR coefficients $\widehat{\Gamma}$.

Network with q nodes. Each node corresponds with a time series.

- draw an **edge** from node i to node j if

$$\sum_{p=1}^{P} |\widehat{\Gamma}_{p,ji}| \neq 0$$

Additionally (if $p = 1$)

- the edge **width** is the size of the effect
- the edge **color** is the sign of the effect (blue if positive, red if negative)

References

- Hsu, Hung, and Chang (2008), "Subset selection for vector autoregressive processes using lasso," *Computational Statistics and Data Analysis*.
- Rothman, Levina, and Zhu (2010), "Sparse multivariate regression with covariance estimation," *Journal of Computational and Graphical Statistics*.
- Basu and Michailidis (2015), "Regularized estimation in sparse high-dimensional time series models," *Annals of Statistics*.
- Gelper S., Wilms I. and Croux C. (2016), "Identifying demand effects in a large network of product categories," *Journal of Retailing*

What about Bayesian statistics?

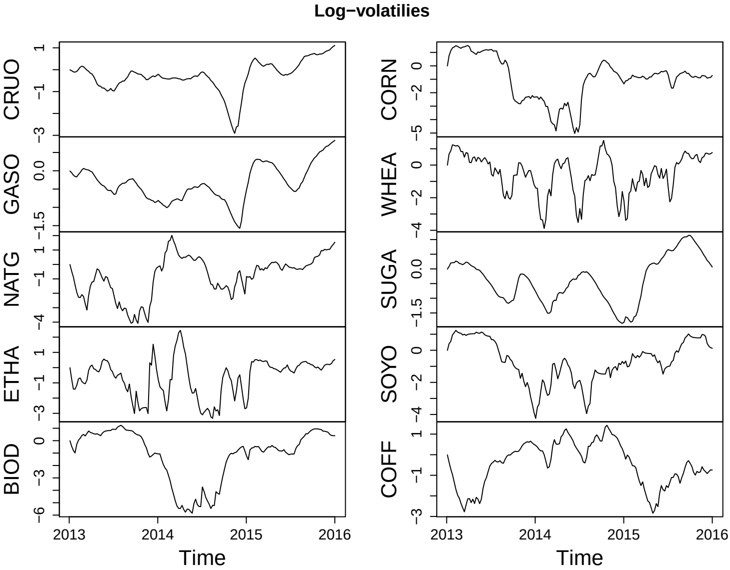
- Bayesian methods
  - Minnesota prior (Koop and Korobilis, 2009)
  - Normal-Inverse Wishart prior (Banbura et al, 2010)

Simulation Design: Sparse high-dimensional : $q = 10, p = 2, T = 50$

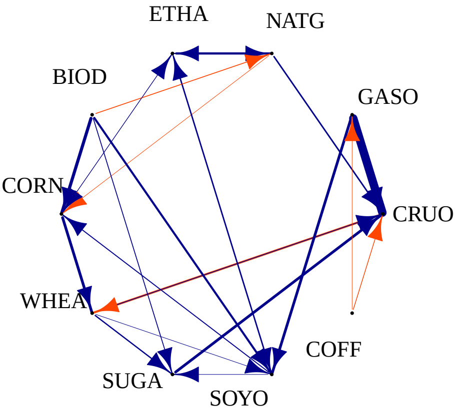
Simulation Study: Results

Method	Mean Absolute Estimation Error
Sparse	0.041
Bayesian: Minnesota	0.044
Bayesian: Normal-Inverse Wishart	0.077
Least Squares	0.157
Restricted LS	0.121

Commodity prices: log volatilities (weekly data)



Network on the AutoRegressive coefficients



Diebold and Yilmaz (2015)

Granger Causality

Time series i is Granger Causing time series j

$$\updownarrow$$

Time series i it has *incremental* predictive power in forecasting series j

$$\updownarrow$$

In the network there is an arrow going from node i to node j

Granger Causality test in high dimensions: Wilms, Gelper, Croux, 2016

Network on the precision matrix