Sparse VAR-Models

Christophe Croux

EDHEC Business School (France)

Joint Work with Ines Wilms (Cornell University), Luca Barbaglia (KU leuven), and Sarah Gelper (TU Eindhoven).
Section 1

VAR Models
10-dimensional time series

Log–volatilities

Time

CRUO
GASO
NA TG
ETHA
BIOD
CORN
WHEA
SUGA
SOYO
COFF

2013 2014 2015 2016

2013 2014 2015 2016

Christophe Croux
Sparse VAR-Models
Two stationary time series $y_{1,t}$ and $y_{2,t}$.

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

\[
\begin{align*}
    y_{1,t} &= \Gamma_{1,11} y_{1,t-1} + \Gamma_{1,12} y_{2,t-1} + e_{1t} \\
    y_{2,t} &= \Gamma_{1,21} y_{1,t-1} + \Gamma_{1,22} y_{2,t-1} + e_{2t}
\end{align*}
\]

Covariance matrix of $(e_{1t}, e_{2t})'$ is $\Sigma$.

Vector notation: $y_t = \Gamma_1 y_{t-1} + e_t$, 
The VAR model

Let $y_t$ be a $q$-dimensional stationary time series

Vector Autoregressive Model of order $p$:

$$y_t = \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \ldots + \Gamma_p y_{t-p} + e_t,$$

- Matrices $\Gamma_j$ are autoregressive parameters
- $e_t$ error with covariance matrix $\Sigma = \Omega^{-1}$
- Standard estimation procedure: OLS equation by equation.
Section 2

Sparse VAR
Log-Volatilies for commodities (weekly data)

VAR model for $q = 10$ time series (and $T = 156$ observations)

- **One lag**
  - $1 \times (q \times q) = 100$ regression parameters
  - 45 unique elements in $\Sigma$

- **Two lags**
  - $2 \times (q \times q) = 200$ regression parameters
  - 45 unique elements in $\Sigma$

→ 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

\[ \hat{\beta} = \arg\min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{l=1}^{k} |\beta_l|. \]

Tibshirani (1996)
Lasso for the VAR model

- Multiple equations
  - Partial correlation structure of the error term
    $\rightarrow$ Glasso of Friedman et al. (2008)

- Dynamic nature of the model
  - Selecting a time series into one of the equations $= \text{selecting the variable and all its lags}$
    $\rightarrow$ Group lasso (Yuan and Lin, 2006)
Penalized ML estimation

Rewrite the VAR in matrix notation:

\[ Y = Y_L \Gamma + E, \]

where

- \( Y = (y_{p+1}, \ldots, y_T)' \)
- \( Y_L = (X_{p+1}, \ldots, X_T)' \) with \( X_t = (y_{t-1}', \ldots, y_{t-p}')' \)
- \( \Gamma = (\Gamma_1, \ldots, \Gamma_p)' \)
- \( E = (e_{p+1}, \ldots, e_T)' \).
Penalized ML estimation (cont.)

Penalized negative log likelihood:

\[
(\hat{\Gamma}, \hat{\Omega}) = \arg\min_{\Gamma, \Omega} \quad \frac{1}{T} \text{tr} \left( (Y - Y_L \Gamma) \Omega (Y - Y_L \Gamma)' \right) - \log |\Omega| \\
+ \lambda_1 \sum_{g=1}^{G} ||\gamma_g||_2 + \lambda_2 \sum_{k \neq k'} |\Omega_{kk'}|,
\]

with

- $\gamma_g$ is subvector of $\Gamma$
- $G = q^2$ total number of groups.
Algorithm

Solving for $\hat{\Gamma} | \Omega$:

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

→ groupwise lasso
Solving for $\Omega|\Gamma$:

$$\hat{\Omega}|\Gamma = \arg\min_{\Omega} \frac{1}{T} \text{tr}\left((Y - Y_L\Gamma)\Omega(Y - Y_L\Gamma)^\prime\right) - \log|\Omega| + \lambda_2 \sum_{k \neq k'} |\Omega_{kk'}|.$$ 

$\rightarrow$ penalized inverse covariance estimation (glasso)
Selection of tuning parameters

In the iteration step $\Gamma|\Omega$, select $\lambda_1$ to minimize

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

- $L_{\lambda_1}$ is the estimated likelihood using $\lambda_1$
- $k_{\lambda_1}$ is the number of non-zero estimated regression coefficients.

In the iteration step $\Omega|\Gamma$, select $\lambda_2$ analogously.
Commodity prices: log volatilities (weekly data)

Log–volatilities

- CRUO
- GASO
- NATG
- ETHA
- BIOD
- CORN
- WHEA
- SUGA
- SOYO
- COFF

Time

2013 2014 2015 2016
Networks from the VAR coefficients $\hat{\Gamma}$.

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

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

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

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)
Network on the AutoRegressive coefficients

Diebold and Yilmaz (2015)
Granger Causality

Time series $i$ is Granger Causing time series $j$

\[\uparrow\]

Time series $i$ it has \textit{incremental} predictive power in forecasting series $j$

\[\downarrow\]

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

Granger Causality \textit{test in high dimensions}: Wilms, Gelper, Croux, 2016
Network on the precision matrix


... 

Section 3

Application: a Market Response Model
Data

Sales, promotion and prices for 17 product categories.

\[ T = 77 \] weekly observations

VAR model for \( q = 3 \times 17 = 51 \) time series
Sales Forecasting

One-step ahead, using rolling window.
Averaged over the 17 product categories, *and the 15 stores*.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Absolute Forecast Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse</td>
<td>736</td>
</tr>
<tr>
<td>Bayesian: Minnesota</td>
<td>875</td>
</tr>
<tr>
<td>Bayesian: Normal-Inverse Wishart</td>
<td>1078</td>
</tr>
<tr>
<td>Least Squares</td>
<td>1298</td>
</tr>
<tr>
<td>Restricted LS</td>
<td>784</td>
</tr>
</tbody>
</table>

Sparse method significantly better ($P < 0.01$, Diebold-Mariano test)
Simulation Study

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

Simulation Design: Sparse high-dimensional: $q = 10, p = 2, T = 50$
<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Absolute Estimation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse</td>
<td>0.041</td>
</tr>
<tr>
<td>Bayesian: Minnesota</td>
<td>0.044</td>
</tr>
<tr>
<td>Bayesian: Normal-Inverse Wishart</td>
<td>0.077</td>
</tr>
<tr>
<td>Least Squares</td>
<td>0.157</td>
</tr>
<tr>
<td>Restricted LS</td>
<td>0.121</td>
</tr>
</tbody>
</table>
Section 4

Extensions
Multi-Class VAR

We have a chain of 17 stores.

- Estimate a VAR for each store separately.
- Pool the data for the 17 stores in one single class, and estimate one single VAR (or panel-VAR.)
- Multi-class estimation, where the data decide what the classes are.

Tibshirani, Saunders, Rosset, Zhu, and Knight (2005): Fused Lasso
Multi-class sparse VAR: $K$ classes

$\Gamma^{(k)}, \ k = 1, \ldots, K$ are minimizing

\[
\sum_{k=1}^{K} \sum_{t=1}^{T} (y_t^{(k)} - \Gamma^{(k)} Y_L^{(k)})' \Omega^{(k)} (y_t^{(k)} - \Gamma^{(k)} Y_L^{(k)}) - TJ \log |\Omega^{(k)}| \\
+ \sum_{k=1}^{K} \sum_{i,j=1}^{J} \sum_{p=1}^{P} |\Gamma^{(k)}_{p,ij}| + \sum_{k \neq k'}^{K} \sum_{i,j=1}^{J} \sum_{p=1}^{P} |\Gamma^{(k)}_{p,ij} - \Gamma^{(k')}_{p,ij}| + \ldots
\]

Wilms, I.; Barbaglia, L.; Croux, C. (2018), JRSSS-C
Furtermore

- t-Lasso for estimating VAR
- cointegration