Sparse High-Dimensional Multivariate Autoregressive Models

Daniel Felix Ahelegbey

University of Venice, Italy
Vrije Universiteit Amsterdam, The Netherlands

July 2, 2014

ISF-Conference 2014, Rotterdam, The Netherlands
Scope and Contributions

★ Propose a Bayesian graphical approach to address:
  ✓ Over-parameterization in MAR models
  ✓ Relationship identification in multivariate time series

★ Methodological Contributions
  ★ Applied an efficient MCMC algorithm for:
    ✓ Joint inference of lag order and temporal dependence in the data
      (with sparsity enforcing constraint on dependence structure)
    ✓ Estimate a predictive MAR (referred to as BGMAR)

★ Application Contributions
  ✓ Modeling and Forecasting Macroeconomic Time Series
Multivariate Autoregressive (MAR) Models

★ MAR models extend the idea of Autoregressive (AR) models for univariate time series to multiple time series - current values of variables are modeled as a linear sum of previous activities

★ MAR models are distinct from structural regression models (e.g. SVAR, SEM) that quantify instantaneous dependence among response variables

★ In a MAR, the variables of interest $Y_t$, is determined by

$$Y_t = \sum_{i=1}^{p} B_i Y_{t-i} + \sum_{i=1}^{p} C_i Z_{t-i} + \epsilon_t$$  \hspace{1cm} (1)
Problems with MAR Estimation

★ Over-parametrization

✓ Parameter Shrinkage (e.g Minnesota prior, Doan et al., 1984)
✓ Variables Selection (e.g SSVS, George et al., 2008)
✓ Parameter Shrinkage and Variables Selection (e.g Lasso, Tibshirani (1996))

★ Relationship Identification

✓ Granger Causality (Granger, 1969)
✓ Lasso-Granger (Arnold et al, 2007 )
Limitation of Standard Approaches

☆ **Over-parametrization**

✓ The Minnesota prior has proved efficient in handling over-parametrized models, its effectiveness is limited

✓ The SSVS is efficient in small dimension models.
  ✫ Parsimony is not guaranteed as unimportant variables are not ignored
  ✫ Computationally intensive for moderate and high dimensional models

✓ Lasso shrinks coefficients of important and unimportant variables at the same rate, over-select the lag order and large size models

☆ **Relationship Identification**

✓ Granger / Lasso Granger relies on regression which may yield significant coefficients for variables that are neither direct nor indirect causes of the response variable.
Graphical Models

- Statistical models that summarize the marginal and conditional independences among variables using graphs (Brillinger, 1996).
  - Represent the logical implication of relationships (e.g. $P \rightarrow Q \rightarrow R$)
  - Suitable representation of the causal relationships using directed edges

- A graphical model $(G, \theta) \in \mathcal{G}, \Theta$ models the joint distribution of variables
  - $G$ the conditional independence structure among variables
  - $\mathcal{G}$ is the space of graphs
  - $\theta$ is a vector of structural parameters
  - $\Theta$ is the parameter space

- Statistical Inference on Graphical Models
  - Quantitative learning: Given $G$ is known, learn $\theta \in \Theta$ from data
  - Structure learning: Given $G$ is unknown, learn $G \in \mathcal{G}$

- This paper focus on both structure and quantitative learning.
Graphical Models and MAR

Let \( X_t = (Y_t, Z_t) = (X^1_t, X^2_t, \ldots, X^n_t) \), \( B = (B(p), C(p)) \)

\[
Y_t = \sum_{s=1}^{p} B_s X_{t-s} + \epsilon_t, \quad X^j_{t-s} \rightarrow Y^i_t \iff B_{s,ij} \neq 0
\]

We define the following notations:

\[
B_s = (G_s \circ \Phi_s), \quad 1 \leq s \leq p
\]  

\( G_s \) is connectivity matrix of structural relationships

\( \circ \) is element-by-element product, (i.e \( B_{s,ij} = G_{s,ij} \Phi_{s,ij} \))

\( B_s \) and \( \Phi_s \) are structural coefficients matrices

\[
B_{s,ij} = \Phi_{s,ij} \quad \text{if} \quad G_{s,ij} = 1, \quad B_{s,ij} = 0 \quad \text{if} \quad G_{s,ij} = 0
\]

The Graphical MAR can be expressed as:

\[
Y_t = \sum_{i=1}^{p} (G_i \circ \Phi_i) X_{t-1} + \epsilon_t
\]
The Bayesian Approaches (Structure Learning)

- Gaussian Likelihood (let $\Omega_x = \Sigma_x^{-1}$)

$$P(X|\Omega_x, G, p) = (2\pi)^{-\frac{nT}{2}} |\Omega_x|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \langle \Omega_x \hat{S}_x \rangle \right\}, \quad \hat{S}_x = \sum_{t=1}^{T} X_t X'_t$$

- Prior on $\Omega_x$ (conditional on complete $G$ and $p$) is Wishart with pdf

$$P(\Omega_x|G, p) = \frac{1}{K_n(\nu, S_0)} |\Omega_x|^{(\nu-n-1)/2} \exp \left\{ -\frac{1}{2} \langle \Omega_x, S_0 \rangle \right\}$$

$$K_n(\nu, S_0) = \int |\Omega_x|^{(\nu-n-1)/2} \exp \left\{ -\frac{1}{2} \langle \Omega_x, S_0 \rangle \right\} d\Omega_x$$

- Marginal Likelihood over graphs and lag order

$$P(X|G, p) = \int P(X|\Omega_x, G, p)P(\Omega_x|G, p)d\Omega_x = (2\pi)^{-\frac{nT}{2}} \frac{K_n(\nu + T, S_0 + \hat{S}_x)}{K_n(\nu, S_0)}$$

- Lag and structure inference

$$(p^*, G^*) = \arg \max_{p, G} P(p, G|X)$$
Efficient Model Inference Scheme

✓ Based on the marginal likelihood function over graphs and lag order

✓ Apply Blocked and Collapsed Gibbs;  
  ( Roberts and Sahu (1997), Liu (1994) )

★ Sample jointly the lag order and graph structure: \((p, G_+ | X)\),  
  \(G_+ = (G_1, G_2, \ldots, G_p)\)

★ Estimate parameters: \((B_+, \Sigma_\varepsilon | p, G_+, X)\),  
  \(B_+ = (B_1, B_2, \ldots, B_p)\)
Apply a reversible jump Markov Chain Monte Carlo (Green (1995))

✓ A birth move: $p \rightarrow p + 1$, $G \rightarrow G^+$, $G$, $n_y \times np$, $G^+$, $n_y \times n(p + 1)$
✓ A death move: $p \rightarrow p - 1$, $G \rightarrow G^-$, $G$, $n_y \times np$, $G^+$, $n_y \times n(p - 1)$
✓ Update move: no change in $p$, $G \rightarrow G'$, $G$ and $G'$ are $n_y \times np$

For lag order (assume truncated Poisson over $p$ with mean ($\lambda = 1$))

For graph structure (assume a uniform prior over $G$ ($P(G|p) \propto 1$))

- we propose a multi-move sampler over $G$
- proposal distribution generates block of indicators (add/delete edge)
- accounting for potential interactions among explanatory variables
- apply the idea of nested graphical model

Apply sparsity enforcing constraint

Select $(G, p)$ that minimize a modified BIC

$$BIC(G, p) = -2 \log(P(X|G, p)) + k \log(T - p), \quad k = \left(\sum_{i=1}^{n_y} |\pi_i| + 2n_y np\right)$$

- $|\pi_i|$ is cardinality of the set of explanatory variables of response variable $Y^i$
Apply Minnesota (MP) / Normal-Wishart (NW) prior distribution

\[ \mathbf{B}_+ \sim \mathcal{N}(\mathbf{B}, \mathbf{V}) \]

Conditional on \( \mathbf{G}_+ \), select relevant variables \( (\mathbf{W}_i) \) for each equation

The posterior mean \( (\overline{\mathbf{B}}_i) \) and variance \( (\overline{\mathbf{V}}_i) \)

\[
\overline{\mathbf{B}}_i = \overline{\mathbf{V}}_i(\overline{\mathbf{V}}_i^{-1}\mathbf{B}_i + \overline{\sigma}_i^{-2}\mathbf{W}_i'\mathbf{Y}_i), \quad \overline{\mathbf{V}}_i = (\overline{\mathbf{V}}_i^{-1} + \overline{\sigma}_i^{-2}\mathbf{W}_i'\mathbf{W}_i)^{-1}
\] (4)

\( \mathbf{B}_i \) and \( \mathbf{V}_i \), are the prior mean and variance of the relevant variables in \( i \)

Under the MP, \( \overline{\sigma}_i = \sigma_i \), \( \Sigma_\varepsilon = \text{diag}(\sigma_1^2, \ldots, \sigma_{n_y}^2) \) assumed to be diagonal, fixed and known

Under NW, \( \overline{\sigma}_i^2 \), is the variance of residuals from the posterior of \( \Sigma_\varepsilon \), assumed to be inverse-Wishart distributed
Modeling and Forecasting Macroeconomic Time Series

★ Moderate: 25-macroeconomic variables from 1959Q1 – 2010Q2, (Koop and Korobilis, 2013),

★ Large: 135-macroeconomic variables from 1959Q1 – 2010Q2

★ 8 - Dependent Variables ($Y_t$)

✓ Gross domestic product, Consumer price index, Federal funds rate, Money stock - M2, Personal consumption, Industrial production index, Unemployment rate, Investment

★ Moderate Dimensional Model : 17 Predictor Variables ($Z_t$)

★ Large Dimensional Model : 127 Predictor Variables ($Z_t$)

★ Compare BGMAR with Lasso

★ Stationarize the series following (Koop and Korobilis, 2013).

★ Maximum lag order $p_{\text{max}} = 4$

★ We use a moving window with an initial sample 1960Q1 – 1974Q4

★ Forecast the next four quarters using a step-ahead forecast.
Results: Moderate Dimensional Model

(a) BIC Score of Graph

(b) Log Score of Model

(c) Number of Links

(d) AIC, Model Predictive Score
Results: Large Dimensional Model

(e) BIC Score of Graph

(f) Log Score of Model

(g) Number of Links

(h) AIC, Model Predictive Score
Results: Moderate vs Large Dimensional Model

(i) Log Score of Model

(j) Number of Links

(k) AIC, Model Predictive Score
✓ Estimate MAR dynamics from the observed time series by
  ✓ Joint inference of lag order and temporal dependence
  ✓ Use graph structure to select relevant variables to estimate MAR models
  ✓ Able to achieve dimension reduction for over-parametrized MAR models

✓ Macroeconomic Application

  ✓ Our inference produce a structure comparable to the Lasso
  ✓ Results shows that Lasso over selects number of edges in high dimensional models
  ✓ Predictive accuracy (AIC) strongly favor the BGMAR over the Lasso in high dimensional settings
  ✓ Key macro indicators can be predicted with a handful of variables
✓ **For Early Warning**

✓ **Evaluating the Interconnectedness of the Financial System**

★ Challenge: Non-existence of fundamental theory on relationship to expect
★ The BIC of the relationship identification strongly favor our approach as a better representation of the structural relationships than pairwise-Granger causality and Lasso
★ Overall, we identified a strongly connected system with linkage between financial and non-financial sectors during major crisis periods

✓ **Further Applications**

★ Augmenting Large VAR with Systemic Risk indicators
★ To model interactions between Macroeconomic variables and Systemic Risk indicators
Thank you! for your attention