Sparse Estimation of Time Series

Abhirup Mallik, Snigdhansu Chatterjee

University of Minnesota

August 2014
Vector Autoregressive Process

- p dimensional vector at time t: $X_t = \{X_{t(1)}, \cdots, X_{t(p)}\}$
- VAR(d) assumes $X_t$ depends on previous d points.

$$X_t = B_1 X_{(t-1)} + \cdots + B_d X_{(t-d)} + \epsilon_t \text{ where } \epsilon_t \sim N_p(0, \Sigma_{\epsilon})$$

- Stationarity: $det(I_p - \sum_{t=1}^{d} B_t z^t) \neq 0$ for all $z \in \mathbb{C}, |z| \leq 1$
- VAR(1) is stationary if Eigen values of $B_1$ have modulus less than 1.
Reformulating VAR(d)

- Var(d) process can be written as a dp dimensional VAR(1) process:
  \[ \tilde{X}_t = \tilde{B}\tilde{X}_{(t-1)} + \tilde{\epsilon} \]

- \[ \tilde{X}_t = \begin{pmatrix} X(t) \\ X(t-1) \\ \vdots \\ X(t-d+1) \end{pmatrix}_{dp \times 1} \]
  \[ \tilde{B} = \begin{pmatrix} B_1 & B_2 & \cdots & B_{d-1} & B_d \\ I_p & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_p & 0 \end{pmatrix}_{dp \times dp} \]

- \[ \tilde{\epsilon} = \begin{pmatrix} \epsilon_t \\ 0 \\ \cdots \\ 0 \end{pmatrix}_{dp \times 1} \]
Graphical Structures

- The time variate relationship between coordinates can be visualized graphically.
- Also possible to correlate multiple time series.
Yule Walker Equations

- Moments of Lag $k$: $\Gamma_k = EX_{t-k}X_t^T$.
- Under Stationarity: $X_t \sim N(0, \Gamma_0)$.

\[
\Gamma_1^T = EX_tX_{t-1}^T \\
= E\{E(BX_{t-1} + e_t)X_{t-1}^T|X_{t-1}\} \\
= BEX_{t-1}X_{t-1}^T \\
= B\Gamma_0^T \\
\Gamma_1 = \Gamma_0B^T
\]
Sparse Estimation of Time Series

Problem formulation

- Objective function based on sample moment equations:
  \[
  \arg \min \| \text{vec}(\hat{\Gamma}_1) - (I_p \otimes \hat{\Gamma}_0)\text{vec}(B^T) \|_2^2 + \lambda \| \text{vec}(B^T) \|_1
  \]
- We propose the following column-wise regularized version of the objective function:
  \[
  \arg \min \frac{1}{p^2} \sum_{i=1}^p \| \Gamma_1^i - \Gamma_0^i B_i^T \|_2^2 + \frac{\lambda}{p} \sum_{i=1}^p \| B_i^T \|_1
  \]
- Column-wise Lasso, parallelizable.
Precision estimation

In the case of single time series, we can also look at the graphical structure of precision matrix as that indicates a dependence structure.

- Residuals: $\hat{R} = (X_t - X_{t-1}\hat{B})$.
- Residual Covariance: $\hat{S} = \frac{1}{T-1}\hat{R}^T\hat{R}$.
- We use Graphical Lasso, for an estimate of $\hat{\Theta} = \hat{\Sigma}^{-1}_\epsilon$ from: $\arg\min_{\Theta} \log \det \Theta - tr(\hat{S}\Theta) - \rho||\Theta||$
Numerical Results

- Simulation Setup: Generated data from VAR(1) with $T=51$, $p=10$. Both transition and error covariance matrix are chosen diagonal.
- Tuning parameters are chosen to minimize the mean squared prediction errors.
Figure: True and estimated transition matrix and precision matrices.
Comparison with Envelope Estimation

- Envelope models try to decompose error variability into two parts: "Material" and "Immaterial" information.
- $\Sigma = \Gamma \Omega \Gamma^T + \Gamma_0 \Omega_0 \Gamma_0^T$
- Where, $\Gamma \in \mathbb{R}^{p \times u}$ and $\Gamma_0 \in \mathbb{R}^{p \times p-u}$ are semi orthogonal basis matrices of $E_\Sigma(B)$ and its orthogonal complement.
- Together they span $\mathbb{R}^p$. And $B = \text{span}(B)$.
- Envelopes are the smallest reduction subspace of $\Sigma$ containing the span of $B$. 
Results

- Error in transition matrix estimation $= \| \hat{B} - B \|_F / \| B \|_F$
- Envelope dimension $= 6$, chosen by BIC.

<table>
<thead>
<tr>
<th>Method</th>
<th>Transition Error</th>
<th>MSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_1$</td>
<td>0.39</td>
<td>7.87</td>
</tr>
<tr>
<td>LL</td>
<td>0.59</td>
<td>7.12</td>
</tr>
<tr>
<td>OLS</td>
<td>0.67</td>
<td>11.19</td>
</tr>
<tr>
<td>Envelope</td>
<td>\ldots</td>
<td>8.13</td>
</tr>
</tbody>
</table>
Figure: True B was Banded, True Σ had a Autoregressive decay. The Estimated transitions and precisions are shown.
Application to Climate data
Application to Climate data
Comparison with Similar Work

- S. Basu et al. dealt with joint estimation of precision and transition matrix.
- Time variate undirected graphs were considered by Zhou et al. and Kolar et al.
- Our work does not consider combined effect of transition matrix and precision matrix.
- Our method is also applicable for establishing relation between several time series.
Thank You
Acknowledgement:

- This research is partially supported by the National Science Foundation under grant # IIS-1029711 and # SES-0851705.
- This research is partially supported by the Institute on the Environment (IonE), and College of Liberal Arts (CLA).