Separable covariance arrays via the Tucker product - Final

by P. Hoff

Kean Ming Tan

June 4, 2013

International Trade Data set

Yearly change in log trade value (in 2000 dollars): $\mathbf{Y} = \{y_{i,j,k,t}\}$

- $i \in \{1, \ldots, 30\}$ indexes the exporting nation
- $j \in \{1, \ldots, 30\}$ indexes the importing nation
- $k \in \{1, \ldots, 6\}$ indexes the commodity type
- $t \in \{1, \ldots, 10\}$ indexes the year

Interested in modeling the mean $M_{ijk} = \mu_{i,j,k}$ across t measurements

What can we do?

International Trade data set

Interested in the model

$$\mathbf{Y} = \mathbf{M} \circ \mathbf{1}_n + \mathbf{E}$$

- iid error model: $\epsilon_{i,j,k,l} \sim \text{normal}(0, \sigma^2)$
- multivariate error model: $\boldsymbol{\epsilon}_{i,j,k} \sim \text{multivariate normal}(\mathbf{0}, \boldsymbol{\Sigma})$
- matrix-variate error model: $\boldsymbol{\epsilon}_{i,j} \sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$

But all four dimensions are correlated!

$$\mathbf{E} \sim ???$$

Definition outline

- Vectorize
- Inner Product
- Matricization
- Array-matrix Product

Definition - k-mode product of an array

Suppose $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ and $\mathbf{A} \in \mathbb{R}^{m_1 \times m_1}$, the 1-mode product is

$$\mathbf{Z} = \mathbf{Y} \times_1 \mathbf{A} \qquad \text{if and only if} \qquad \mathbf{Z}_{(1)} = \mathbf{A}\mathbf{Y}_{(1)}$$

Array-matrix multiplication: $\mathbf{Y} \times_1 \mathbf{A}$

1. Matricization: $\mathbf{Y}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Y}_{(1)}$
3. Reform: $\mathbf{Y} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Y}_{(1)}), m_1, m_2, m_3)$

Remember this picture?

We managed to find the Array Normal Model

$$\mathbf{Y} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K)$$

Array normal model[1]

Array normal model, $\mathbf{y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$:

$$
\begin{aligned}
\mathbf{Z} = \{z_{i,j,k}\} &\overset{\text{iid}}{\sim} \text{normal}(0,1) \\
\mathbf{Y} &= \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\
&\overset{\text{iid}}{\sim} \text{array normal}(\mathbf{M}, \mathbf{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \mathbf{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \mathbf{\Sigma}_3 = \mathbf{C}\mathbf{C}^T) \\
&\sim \text{array normal}(\mathbf{M}, \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3)
\end{aligned}
$$

[1] Adapted from Hoff's slides

Probability density function

Multivariate normal

$$(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp(-(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2)$$

Array normal: Note that $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3\}$

$$(2\pi)^{-m/2}(\prod_{k=1}^{3}|\boldsymbol{\Sigma}_k|^{-m/(2m_k)})\exp(-\|(\mathbf{Y}-\mathbf{M})\times\boldsymbol{\Sigma}^{-1/2}\|^2/2)$$

where $m = m_1 m_2 m_3$

Estimation: Frequentist

- $\hat{\mathbf{M}} = \bar{\mathbf{Y}} = \sum \mathbf{Y}_i / n$
- Let $\mathbf{E} = \mathbf{Y} - \bar{\mathbf{Y}} \circ \mathbf{1}_n$ and repeat the following for $k = 1, 2, 3$
  - Compute $\tilde{\mathbf{E}} = \mathbf{E} \times \{\boldsymbol{\Sigma}_1^{-1/2}, \mathbf{I}_{m_k}, \boldsymbol{\Sigma}_3^{-1/2}, \mathbf{I}_n\}$ and $\mathbf{S}_k = \tilde{\mathbf{E}}_{(k)} \tilde{\mathbf{E}}_{(k)}^T$
  - $\boldsymbol{\Sigma}_k = \mathbf{S}_k / n_k$ where $n_k = n \times \prod_{j \neq k} m_j$

Analogy for univariate case $\sigma^2 = \sum (y - \bar{y})^2 / n$

Estimation: Bayesian

Assume the following prior:

- $\mathbf{M}|\mathbf{\Sigma} \sim \text{anorm}(\mathbf{M}_0, \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 / \kappa_0)$
- $\mathbf{\Sigma}_k \sim \text{inverse-Wishart}(\mathbf{S}_{0k}^{-1}, v_{0k})$

After A LOT of calculation, the posteriors are

- $\mathbf{M}|\mathbf{Y}, \mathbf{\Sigma} \sim \text{anorm}([\kappa_0 \mathbf{M}_0 + n\bar{\mathbf{Y}}]/[\kappa_0 + n], \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 / [\kappa_0 + n])$
- $\mathbf{\Sigma}_k|\mathbf{Y}, \mathbf{\Sigma}_{-k} \sim \text{inverse-Wishart}([\mathbf{S}_{0k} + \mathbf{S}_k + \mathbf{R}_{(k)}\mathbf{R}_{(k)}^T]^{-1}, v_{0k} + n_k)$

where $\mathbf{R} = \sqrt{\frac{\kappa_0 n}{\kappa_0 + n}}(\bar{\mathbf{Y}} - \mathbf{M}_0) \times \{\mathbf{\Sigma}_1^{-1/2}, \mathbf{I}_{m_k}, \mathbf{\Sigma}_3^{-1/2}\}$

Finally!!! We model this data set!

Want to model

$$\mathbf{Y} = \mathbf{M} \circ \mathbf{1}_n + \mathbf{E}$$

Understand correlation among exporters, importers and commodities of **E**

Consider two models:

1. $\mathbf{Y} \sim \text{anorm}(\mathbf{M} \circ \mathbf{1}_n, \mathbf{I}_{m_1} \circ \mathbf{I}_{m_2} \circ \mathbf{\Sigma}_3 \circ \mathbf{\Sigma}_4)$
2. $\mathbf{Y} \sim \text{anorm}(\mathbf{M} \circ \mathbf{1}_n, \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \circ \mathbf{\Sigma}_4)$

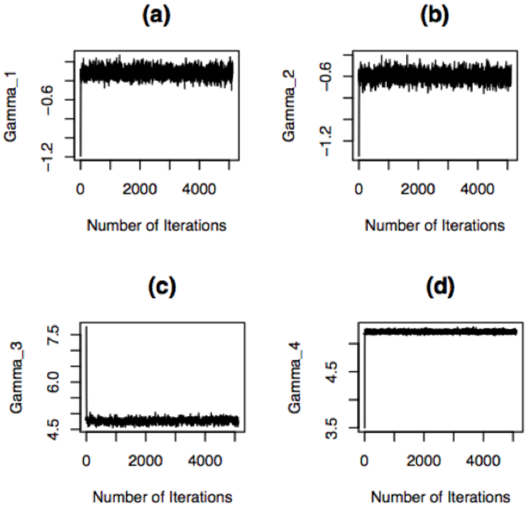
MCMC

- 205,000 iterations, 5,000 as burn-in period
- Parameter values are kept only every 40th iteration
- We have 5,000 samples

But samples are matrices, how to check for convergence?

Jon Wakefield: Calculate a summary measure of the matrix, for instance, the determinant

Trace plots


(a)
Gamma_1
−1.2
−0.6
0
2000
4000
Number of Iterations
(b)
Gamma_2
−1.2
−0.6
0
2000
4000
Number of Iterations
(c)
Gamma_3
4.5
6.0
7.5
0
2000
4000
Number of Iterations
(d)
Gamma_4
3.5
4.5
0
2000
4000
Number of Iterations

Model Comparison

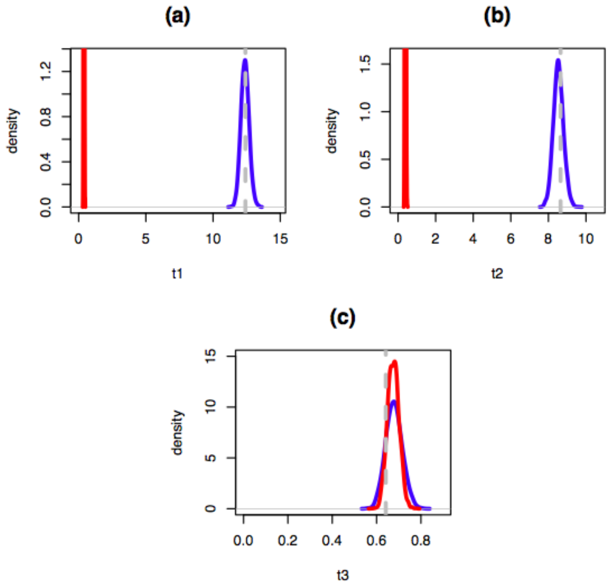
Posterior predictive evaluation:

$$t_1(\mathbf{Y}) = \log|\mathbf{I}/m_1| - \log|\tilde{\mathbf{S}}_1|$$

where $\mathbf{S}_1 = \mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$ and $\tilde{\mathbf{S}}_1 = \mathbf{S}_1/\text{tr}(\mathbf{S}_1)$

Discrepancies between $t_1(\mathbf{Y})$ and $t_1(\tilde{\mathbf{Y}})$ indicates model does not capture some aspects of the data

Results

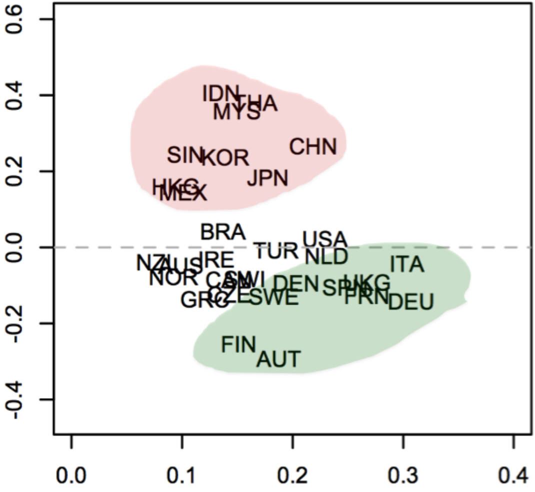


(a)
density
0.0
0.4
0.8
1.2
0
5
10
15
t1
(b)
density
0.0
0.5
1.0
1.5
0
2
4
6
8
10
t2
(c)
density
0
5
10
15
0.0
0.2
0.4
0.6
0.8
t3

Understand Correlation

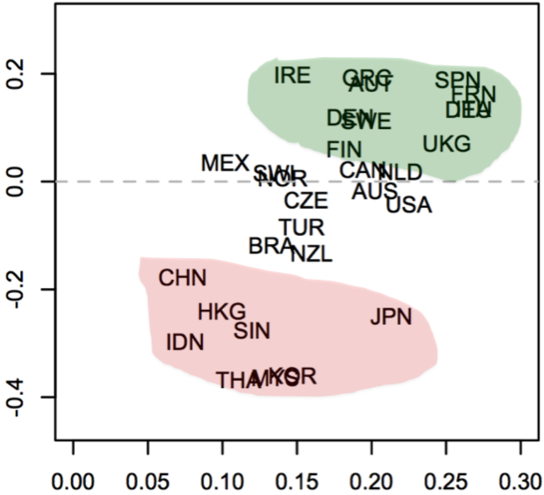
- Understand relationship among exporters, importers, and commodities, respectively
- Analyze posterior mean estimates of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, and $\boldsymbol{\Sigma}_3$
- Plot first two eigenvectors of each estimate

Exporter



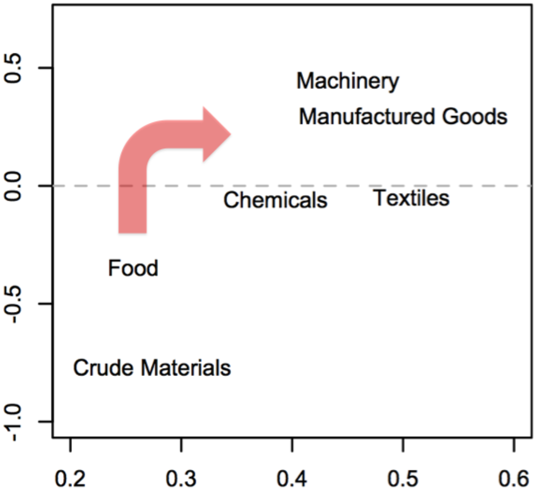
IDN
THA
MYS
CHN
SIN
KOR
JPN
HKG
MEX
BRA
USA
TUR
NLD
NZL
AUS
IRE
ITA
NOR
CAN
SWI
DEN
SPN
HKG
CZE
FRN
GRC
SWE
DEU
FIN
AUT
0.6
0.4
0.2
0.0
-0.2
-0.4
0.0
0.1
0.2
0.3
0.4

Importer



0.2
0.0
-0.2
-0.4
0.00
0.05
0.10
0.15
0.20
0.25
0.30
IRE
GRC
AUT
SPN
FRN
DEN
SWE
FIN
UKG
MEX
NOR
CAN
NLD
AUS
USA
CZE
TUR
BRA
NZL
CHN
HKG
SIN
IDN
JPN
KOR

Commodity


Machinery
Manufactured Goods
Chemicals
Textiles
Food
Crude Materials
0.5
0.0
-0.5
-1.0
0.2
0.3
0.4
0.5
0.6

Discussion & Critique - 1

Estimation of parameters:

- Bayesian approach preferable
- MLE might not exist
- Estimation of high-dimensional parameters benefits from regularization

Discussion & Critique - 1 cont

Impose regularization on the inverse covariance matrix $\boldsymbol{\Sigma}_k^{-1}$

$$\ell(\mathbf{Y}|\boldsymbol{\Sigma}) \propto \sum_{k=1}^{K} \frac{n_k}{2} \log|\boldsymbol{\Sigma}_k^{-1}| - \frac{1}{2} \sum_{i=1}^{n} \|(\mathbf{Y}_i - \bar{\mathbf{Y}}) \times \boldsymbol{\Sigma}^{-1/2}\|^2 - \sum_{k=1}^{K} \lambda_k \|\boldsymbol{\Sigma}_k^{-1}\|_1$$

As usual, $\hat{\mathbf{M}} = \bar{\mathbf{Y}}$
Closed form solution for $\boldsymbol{\Sigma}^{-1} = \{\boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_K^{-1}\}$

DOES NOT EXIST!!!

Discussion & Critique -1 cont

How about an iterative approach? Given $\hat{\mathbf{M}} = \bar{\mathbf{Y}}$

$$\ell(\mathbf{Y}|\boldsymbol{\Sigma}) \propto_{\boldsymbol{\Sigma}_k^{-1}} \frac{n_k}{2}\log|\boldsymbol{\Sigma}_k^{-1}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_k^{-1}\tilde{\mathbf{E}}_{(k)}\tilde{\mathbf{E}}_{(k)}^T\right) - \lambda_k\|\boldsymbol{\Sigma}_k^{-1}\|_1$$

Block coordinate descent approach!

Discussion & Critique - 2

Josh Keller: Isn't normality assumption too stringent for an array?

Discussion & Critique - 2

Josh Keller: Isn't normality assumption too stringent for an array?

Normality assumption

Central Limit Theorem, no worries:)

Discussion & Critique - 2

Josh Keller: Isn't normality assumption too stringent for an array?

Normality assumption

Central Limit Theorem, no worries:)

Not quite true in array case

Some other types of array models are clearly needed!

Discussion & Critique - 2

Working towards array t-distribution

$\mathbf{x} \in \mathbb{R}^p$ has elliptical contoured distribution with mean $\boldsymbol{\mu}$ and dispersion parameter $\boldsymbol{\Sigma}$, call it $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h(\cdot))$

$$f(\mathbf{x}) = c_n |\boldsymbol{\Sigma}|^{-1/2} h\left((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Special Cases:

- Multivariate normal distribution: $h(\mathbf{z}) = \exp(-\frac{1}{2}\mathbf{z})$
- Multivariate t-distribution: Under some transformation and rotation

Discussion & Critique - 2

$\mathbf{X} \in \mathbb{R}^{m_1 \times \cdots \times m_K}$ has array elliptical contoured distribution with mean $\mathbf{M}$ and covariance matrices $\mathbf{\Sigma} = \{\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K\}$, $E_{m_1,\ldots,m_K}(\mathbf{M}, \mathbf{\Sigma}, h(\cdot))$

$$f(\mathbf{X}) = c_n \prod_{k=1}^{K} |\mathbf{\Sigma}_k|^{-m/2m_k} h(\|(\mathbf{Y} - \mathbf{M}) \times \mathbf{\Sigma}^{-1/2})\|^2)$$

Special Cases:

- Array normal distribution: $h(\mathbf{Z}) = \exp(-\frac{1}{2}\mathbf{Z})$
- Array t-distribution: ???

Array t-distribution

Conclusion

- Construct the array normal distribution
- Model the International Trade Data Set

Questions?