Model Selection and Adaptive Markov Chain Monte Carlo for Bayesian Cointegrated VAR Models

Gareth W. Peters\textsuperscript{1}, B. Kannan\textsuperscript{2}, Ben Lasscock\textsuperscript{2} and Chris Mellen\textsuperscript{2}

\textsuperscript{1}CSIRO Mathematical and Information Sciences, NSW
\textsuperscript{2}UNSW School of Mathematics and Statistics, NSW
\textsuperscript{2}Boronia Capital Pte. Ltd., NSW
Overview

- Literature
- Novelty
- Problem formulation
- Adaptive MCMC Algorithms
- Simulation and Results
  - Analysis of proposed sampling methodology on synthetic data sets
  - Bayes Factor analysis on synthetic and real data sets
  - Analysis of predictive performance of BMOS and BMA
- Conclusions
• Bayesian analysis of CVAR – Koop et al. (2006), Reinsel et al. (1998). Priors have to be carefully designed to avoid improper posteriors.

• ECM conditioned on cointegration vectors, $\beta$- Gweke (1995), Sugita(2002).

• Generally griddy Gibbs samplers are used in the Cointegration Literature:
  • Bauwens and Luibrano(1996)
  • Sugita(2002)

• Model order selections via Bayes Factor analysis - Sugita(2002)
Novelty: Adaptive Matric-variate MCMC Methodology in CVAR Models

• Griddy-Gibbs samplers suffer from the “curse of dimensionality” and alternatives such as importance sampling are also problematic in high dimensions.

(unattainable optimal importance sampling distribution $\rightarrow$ High Variance)

• We propose “adaptive matric-variate MCMC methodology” with properties:
  • Doesn’t suffer from the curse of dimensionality
  • Simple to implement
  • Can be automated with varying dimensionality

• With these new samplers, rank, $r$ of a reduced rank Cointegration model is estimated via Bayes Factor (BF) analysis from the posterior distributions.

• We also demonstrate how these posterior distributions can be used under Bayesian Model Averaging (BMA) framework to remove rank uncertainties in predictions and moment estimations.
Adaptative Matric-variate MCMC Methodology

• Distinguishing feature of Adaptative MCMC algorithms is generation of the Markov Chain samples from the target matric-variate posterior distribution via a time and path space dependent adaptive sequence of kernals.

• This is a novel new class of algorithms using combinations of time or state inhomogeneous proposal kernals, where the proposal distributions vary according to the past history of the Markov chains.

• Design is DELICATE in order to preserve the correct matric-variate target posterior distribution as the stationarity distribution of the resulting Markov chain.

• Several theoretical conditions and restrictions discussed in the literature to ensure the ergodicity of the adaptive algorithms.
Roberts and Rosenthal (2005, 2008) have introduced simpler conditions known as “Diminishing Adaptation and Bounded Convergence”:

- **Diminishing Adaptation:**
  \[
  \lim_{n \to \infty} \sup_{\theta \in E} \| Q_{\Gamma_{n+1}}(\theta, \cdot) - Q_{\Gamma_n}(\theta, \cdot) \| = 0 \text{ in probability.}
  \]

- **Bounded Convergence:**
  \[
  \{ M_\varepsilon (\Theta^n, \Gamma_j) \}_{j=0}^\infty \text{ is bounded in probability, } \varepsilon > 0.
  \]

Which guarantee asymptotic convergence in two senses:

- **Asymptotic Convergence:**
  \[
  \lim_{n \to \infty} \| L(\Theta^n) - P(\cdot) \| = 0 \text{ in probability.}
  \]

- **WLLN:**
  \[
  \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n g(\Theta^i) = \int g(\theta) p(\theta) d\theta \text{ for all bounded } g : E \to \mathbb{R}
  \]
Problem Formulation

• An unrestricted VECM representation of a n-dimensional I(1) time series, $X_t$ is given by

$$\Delta X_t = \mu + \alpha \gamma z_t + \sum_{i=1}^{p-1} \Psi_{t-i} \Delta X_{t-i} + \epsilon_t$$

where

$z_t = [X'_{t-1}, t, \gamma = [\beta' \delta], t = p, p+1, \ldots , T,$

$p$ – number of lags in VAR

$\epsilon_t \sim N(0, \Sigma_n)$ and independent over time.

$\mu$ – $n \times 1$ trend coefficients

$\Psi_i$ – $n \times n$ ith autoregressive coefficients

$\alpha$ – $n \times r$ adjustment coefficients

$\beta$ – $n \times r$ cointegration vector

$\delta$ – $r \times 1$ drift coefficients which is zero in this paper
Problem Formulation (Contd...)  

- Above representation can be reformulated into a matrix format as:

\[
\begin{bmatrix}
\Delta X'_{p} \\
\Delta X'_{p+1} \\
\vdots \\
\Delta X'_{T}
\end{bmatrix}
\begin{bmatrix}
1 & \Delta X'_{p-1} & \Lambda & \Delta X'_{1} \\
1 & \Delta X'_{p} & \Lambda & \Delta X'_{2} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \Delta X'_{T-1} & \Lambda & \Delta X'_{T-p+1}
\end{bmatrix}
\begin{bmatrix}
\mu' \\
\Psi_1' \\
\vdots \\
\Psi_p' \\
\gamma' \alpha' \\
\varepsilon_p' \\
\varepsilon_{p+1}' \\
\varepsilon_T'
\end{bmatrix}
\]

\[= X\Gamma + Z\gamma' \alpha' + E = [X \ Z\gamma'] \times \begin{bmatrix} \Gamma \\ \alpha' \end{bmatrix} + E \]

\[= WB + E \]

where \( E \sim N(0_{(T-p+1)\times n}, I_{(T-p+1)} \otimes \Sigma_n) \): matrix- variate normal distribution with \( 0_{(T-p+1)\times n} \) mean and \( I_{(T-p+1)} \otimes \Sigma_n \) covariance matrix.

- Our aim is to estimate \( \alpha, \beta, B, \Gamma \) and the model order, \( r \).
Assumptions and Restrictions

• Identification issue:

\[ \Pi = \alpha \beta' = \alpha AA^{-1} \beta' \]

\[ \rightarrow \text{Imposer}^2 \text{ restrictions as } \beta = [I_r, B*]' \]

• Terminal absorbing state when \( \alpha = 0 \): make sure that the sampler or Markov chain doesn’t enter the terminal state.

• Nonlinear ECM model becomes linear when conditioned on \( \beta \).

• Use conjugate hierarchical priors when possible.
Likelihood function can be derived from the matrix-variate normal distribution of $E$: 

$$L(Y \mid B, \Sigma_n, \beta) = (2\pi)^\frac{-(T-P+1)n}{2} |\Sigma_n|^{\frac{(T-P+1)n}{2}} \text{etr}[-0.5 \times \Sigma_n^{-1}(\hat{S} + R)]$$

where $R = (B - \hat{B})'WW(B - \hat{B})$, $\hat{S} = (Y - W\hat{B})'(Y - W\hat{B})$ and $\hat{B} = (W'W)^{-1}W'Y$
Prior Specifications: Similar to Sugita (2002)

- We use Conjugate hierarchical priors:

\[ \Sigma_n \sim IW_n(h, S) : \text{matrix-variate Inverted Wishart distribution with} \]
\[ h \text{ degree of freedom and PDS, } S(n \times n) = \tau Y'Y \]
\[ B | \Sigma_n \sim N(P, \Sigma_n \otimes A^{-1}) : \text{matrix-variate Gaussian distribution with} \]
\[ \text{mean } P = (\hat{W}'\hat{W})^{-1} \hat{W}Y \text{ and covariance matrix } \Sigma_n \otimes A^{-1}, \text{ where} \]
\[ \Sigma_n = I_n, A(k \times k) = \lambda \frac{\hat{W}'\hat{W}}{T} \text{ where } \hat{W} = XZ\hat{\beta}, \hat{\beta} = [I_r; 0_{n-r}] \text{ and } \lambda = 1 \]
\[ \beta \sim N(\overline{\beta}, Q \otimes H^{-1}) : \text{matrix-variate Gaussian distribution with} \]
\[ \text{mean } \overline{\beta} = [I_r; 0_{n-r}] \text{ and covariance matrix } Q \otimes H^{-1}, \text{ where} \]
\[ Q = I_n \text{ and } H(n \times n) = \tau Z'Z \text{ with } \tau = 1 \]
\[ \text{where } k = 1 + n(p-1) + r \]
Posterior Distributions: Similar to Sugita(2002)

- Posterior distributions (without derivations details):

\[ \Sigma_n \mid \beta, Y \sim IW_n(T - p + 1 + h, S_\star) : \text{matric- variate inverted Wishart distribution} \]
\[ B \mid \beta, Y \sim T_{k,n}(T - p - n + 1, S_\star, A_\star, B_\star) : \text{matric- variate Student t- distribution} \]
\[ \beta \mid Y \propto p(\beta) \mid S_\star \sim (T - p + 1 + h + 1)/2 \mid A_\star \sim n/2 : \text{non-standard distribution} \]

where

\[ A_\star = A + W'W \]
\[ B_\star = (A + W'W)^{-1}(AP + W'WB) \]
\[ S_\star = S + \hat{S} + (P - \hat{B})'[A^{-1} + (W'W)^{-1}]^{-1}(P - \hat{B}) \]
Algorithms: Adaptive MCMC Schemes

• The posterior distribution for $\beta|Y$ is a non-standard distribution, thus one can not use a straight forward inverse sampler. Very little is known about this posterior (complicate the MCMC).

• Instead we propose two alternative samplers using adaptive matric-variate MCMC methodology.

• **Algorithm 1**: Offline adaptively pre-tuned mixture proposal with local and global random walk moves:

\[
q(\tilde{\beta}^{t-1},.) = w_1N(\tilde{\beta}; \tilde{\beta}^{ML}, \Sigma_{n}^{ML}) + (1 - w_1)N(\tilde{\beta}_{i,k}; \tilde{\beta}_{i,k}^{t-1}, \sigma_{i,k}^2)
\]

• **Algorithm 2**: Online matric-variate adaptive Metropolis algorithm discussed in Roberts and Rosenthal(2008), where the proposal distribution at the $j^{th}$ iteration is:

\[
q(\tilde{\beta}^{t-1},.) = w_1N(\tilde{\beta}; \tilde{\beta}^{t-1}, \frac{2.38^2}{d} \Sigma_{n,j}) + (1 - w_1)N(\tilde{\beta}; \tilde{\beta}^{t-1}, \frac{0.1^2}{d} I_d)
\]

where $d = (n-r) \times r$
Rank Estimations in BCVAR Models Using Bayes Factor Analysis

- With conjugate priors on $\alpha$, the rank of the unrestricted $\alpha$ are compared to the 0 rank setting as:
  $$\Pr(r \mid Y) = \frac{BF_{r/0}}{\sum_{j=0}^{n} BF_{j/0}}$$
  where $r = 1, \ldots, n$ and $BF_{0/0} = 1$

- We used the generalized Save-Dickey density ratio of Verdinelli et al. (1995) for nested model structure to calculate $BF_{r/0}$:
  $$BF_{r/0} = \frac{p(\alpha' = 0_{r \times n})}{C_r^{-1} p(\alpha' = 0_{r \times n} \mid Y)} = \frac{\int p(\alpha, \beta, \Gamma, \Sigma_n \mid Y) d\alpha d\beta d\Gamma d\Sigma}{C_r^{-1} \int p(\alpha, \beta, \Gamma, \Sigma_n \mid Y)_{\text{rank}(\alpha) = 0} d\alpha d\beta d\Gamma d\Sigma}$$
  where
  $$C_r = \int p(\alpha, \beta, \Gamma, \Sigma_n)_{\text{rank}(\alpha) = 0} d\beta d\Gamma d\Sigma$$
Simulations

Simulations and analysis are done in three parts:

• Analysis of proposed sampling methodology on synthetic data sets
• Bayes Factor analysis on synthetic and real data sets
• Analysis of predictive performance of BMOS and BMA

Generic Simulations Parameters:

• \( T = 100 \)
• Number of realizations = 20
• Number of samples \( J = 20,000 \)
• Burn-in samples = 10,000
• \( n = 4 \)
Analysis of Proposed Sampling Methodology (synthetic data)

- For a rank 2 model, we have estimated average MMSE and average posterior standard deviations for $\beta$, $\alpha$, and $\text{trace}(\Sigma_n)$, which are presented in tables 1-3 respectively.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Algo. 1</th>
<th>Algo. 2</th>
<th>True val</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{MMSE } B_{1,r+1}$</td>
<td>-0.002(0.001)</td>
<td>-0.034(0.002)</td>
<td>0</td>
</tr>
<tr>
<td>Posterior Stdev $B_{1,r+1}$</td>
<td>0.018(0.006)</td>
<td>0.010(0.003)</td>
<td>-</td>
</tr>
<tr>
<td>$\text{MMSE } B_{2,r+1}$</td>
<td>-0.819(0.051)</td>
<td>-0.862(0.045)</td>
<td>-1</td>
</tr>
<tr>
<td>Posterior Stdev $B_{2,r+1}$</td>
<td>0.032(0.005)</td>
<td>0.020(0.003)</td>
<td>-</td>
</tr>
<tr>
<td>$\text{MMSE } B_{1,n}$</td>
<td>0.033(0.025)</td>
<td>-0.024(0.023)</td>
<td>0</td>
</tr>
<tr>
<td>Posterior Stdev $B_{1,n}$</td>
<td>0.030(0.012)</td>
<td>0.026(0.010)</td>
<td>-</td>
</tr>
<tr>
<td>$\text{MMSE } B_{2,n}$</td>
<td>-0.752(0.098)</td>
<td>-0.774(0.082)</td>
<td>-1</td>
</tr>
<tr>
<td>Posterior Stdev $B_{2,n}$</td>
<td>0.038(0.013)</td>
<td>0.028(0.006)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table (1): Estimates of $\beta$

- Average mean acceptance Probs. for $\beta$ are 0.352 and 0.232 in algorithm 1 and algorithm 2, respectively.
• Estimates of $\alpha$:

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Algo. 1</th>
<th>Algo. 2</th>
<th>True val</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE $\alpha_{1,1}$</td>
<td>-0.223(0.015)</td>
<td>-0.224(0.016)</td>
<td>-0.2</td>
</tr>
<tr>
<td>Posterior Stdev. $\alpha_{1,1}$</td>
<td>0.070(0.006)</td>
<td>0.068(0.005)</td>
<td>-</td>
</tr>
<tr>
<td>MMSE $\alpha_{1,2}$</td>
<td>0.201(0.013)</td>
<td>0.202(0.013)</td>
<td>0.2</td>
</tr>
<tr>
<td>Posterior Stdev. $\alpha_{1,2}$</td>
<td>0.053(0.002)</td>
<td>0.052(0.002)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table (2): Estimates of $\alpha$
• Estimates of \( \text{trace}(\Sigma_n) \)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Algo. 1</th>
<th>Algo. 2</th>
<th>True val</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE trace(( \Sigma_n ))</td>
<td>4.945(0.331)</td>
<td>4.432(0.332)</td>
<td>4</td>
</tr>
<tr>
<td>Posterior Stdev. trace(( \Sigma_n ))</td>
<td>0.420(0.049)</td>
<td>0.416(0.048)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table (3): Estimates of trace(\( \Sigma_n \))

• These results in the above table demonstrates that both adaptive MCMC algorithms estimate the interested parameters accurately.

• In addition, we point out that these samplers actually achieve optimal performance as \( d \) approaches infinity - [see Rosenthal (2008)]
Convergence Properties of the Proposed Algorithms (synthetic data)

• We have plotted one realization of the Markov chains for $\beta$, $\text{trace}(\Sigma)$ and $B$ in figure(1) below:

![Markov Chains](image)

Figure(1): Markov chains for $\beta$, $\text{trace}(\Sigma)$ and $B$

• Above figures shows that Markov chains converges very quickly to the true values, even when the starting points are far away from the true values.

• Additionally, the adaptive nature of the proposal kernels is clearly evident as the proposal learns online the appropriate proposal variance, thus affecting the mixing of the chains.
Rank Estimations in BCVAR Models Using Bayes Factor Analysis (synthetic data)

- One needs to take care of numerical complications that arise when calculating Savage-Dickey density ratio.
- Specifically when the number of samples increases the term \( |S_{**}^{i} k+h|^{2} \) will explode numerically.
- Appendix 1 of our paper details steps critical to avoid such numerical issues.
- Simulation results when the true model orders are 1 and 3 are shown below in tables 4 and 5 respectively. We use \( T = 100 \), and the MAP estimates of \( r \) were made from 20 independent data sets.

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Bayes Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3 (0.84)</td>
</tr>
<tr>
<td>1</td>
<td>16 (0.93)</td>
</tr>
<tr>
<td>2</td>
<td>2 (0.92)</td>
</tr>
<tr>
<td>3</td>
<td>0 (-)</td>
</tr>
<tr>
<td>4</td>
<td>0 (-)</td>
</tr>
</tbody>
</table>

Table 4: true rank = 1

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Bayes Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 (-)</td>
</tr>
<tr>
<td>1</td>
<td>0 (-)</td>
</tr>
<tr>
<td>2</td>
<td>4 (0.89)</td>
</tr>
<tr>
<td>3</td>
<td>6 (0.90)</td>
</tr>
<tr>
<td>4</td>
<td>10 (0.94)</td>
</tr>
</tbody>
</table>

Table 5: true rank = 13
• With the proposed samplers, BF methodology is clearly able to detect the true model order in high proportions.

• Averaged posterior model probabilities are very selective of the correct model, indicating there won’t be great benefit, at least under synthetic data scenarios, in performing model averaging.

• There will be a benefit in model averaging whenever the posterior model (rank) probability is spread over more than one likely model.

We found this to be the case in two scenarios:

• small data sets $T < 50$

• if the actual rank changes throughout the series
Rank Estimations in BCVAR Models Using Bayes Factor Analysis (real data)

- Simulations set-up:
  - Markets:
    - US indexes: S&P mini, Nasdaq mini and Dow Jones mini
    - US notes: 5 Year, 10 Year and 30 Year notes
  - Data: Daily closing prices, number of samples varied from 50 to 350 (in 50 samples increment)
  - Twenty independent realizations for each possible ranks with different initializations.
  - Length of Markov chains: 20,000
• Results of BF analysis are presented in tables 6 and 7 for the US indexes and notes, respectively.

<table>
<thead>
<tr>
<th>Table(6): Log BFs: US indexes</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>r \ T</th>
<th>50</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>8.09(0.78)</td>
<td>3.77(0.29)</td>
<td>7.01(0.5)</td>
<td>11.51(0.9)</td>
<td>1.69(0.5)</td>
<td>2.71(3.6)</td>
<td>3.1(1.1)</td>
</tr>
<tr>
<td>2</td>
<td>2.91(1.24)</td>
<td>2.33(1.26)</td>
<td>4.61(0.63)</td>
<td>25.4(7.2)</td>
<td>-5.3(1.2)</td>
<td>-5.8(1)</td>
<td>4.9(1.1)</td>
</tr>
<tr>
<td>3</td>
<td>-26.03(1.6)</td>
<td>-8.45(0.27)</td>
<td>-37.25(1.08)</td>
<td>-55.8(1.7)</td>
<td>-14.6(0.03)</td>
<td>-62.6(3.3)</td>
<td>8.9e-3(0.88e-3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table(7): Log BFs: US indexes</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>r \ T</th>
<th>50</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>4.81(1)</td>
<td>3.1(0.4)</td>
<td>5.4(1.1)</td>
<td>5.9(0.8)</td>
<td>3.3(0.8)</td>
<td>1.3(0.4)</td>
<td>7.3(0.6)</td>
</tr>
<tr>
<td>2</td>
<td>-1.67(12.8)</td>
<td>3.7(3.9)</td>
<td>-3.8(3)</td>
<td>-1.8(2.6)</td>
<td>-6(3.2)</td>
<td>-0.14(6.5)</td>
<td>-2.9(1.96)</td>
</tr>
<tr>
<td>3</td>
<td>-42.4(12.4)</td>
<td>-48.65(2.85)</td>
<td>-32.1(0.1)</td>
<td>-100.44.8</td>
<td>-25.9(0.06)</td>
<td>-10.3(0.7)</td>
<td>-142.9-3(3.3)</td>
</tr>
</tbody>
</table>
• From the above tables, one can see that predicted rank varies with the sample size for US indexes.

• The model gives preference to rank 1 suggesting that two common trends are present in the series and in several cases.

• For the US indexes it is less likely to distinguish between rank 1 and 2: Model Averaging may help??

• For the US notes, the results in table(7) shows that there is stronger evidence for a single cointegration relationship over time compared to the US indexes.
Predictive Performance Comparison: BMOS vs BMA

• In BMOS, most probable posterior model corresponding to MAP, \( r_{MAP} \) is selected and samples \( \{\alpha, \beta, B, \Sigma\} \) [or functions of samples \( \varphi(\{\alpha, \beta, B, \Sigma\}) \)] from the corresponding Markov chain are used in the parameter estimation or moment calculations.

• However, when there are ambiguities in distinguishing between models, Bayesian Model Averaging (BMA) can remove the model risk associated with the rank uncertainty.

• For example, one can perform prediction under BMA as:

\[
Pr(Y^* \mid Y) = \sum_r \int p(Y^* \mid \beta, B, \Sigma_n, r) p(\beta, B, \Sigma_n \mid Y, r) p(r \mid Y) d\beta dB d\Sigma_n
\]

• The BMOS counterpart of such prediction involves:

\[
Pr(Y^* \mid Y) = \int p(Y^* \mid \beta, B, \Sigma_n, r_{MAP}) p(\beta, B, \Sigma_n \mid Y, r_{MAP}) d\beta dB d\Sigma_n
\]
Predictive Performance Comparison: BMOS vs BMA

- We use the US 5 Yr, 10 Yr notes and S&P 500 mini in analyzing the prediction performance.
- We compare the MMSE estimates of the predicted series over 10 steps ahead after removing the rank uncertainties under BMA against the BMOS counterpart.
- In Figure (2), below we present a histogram of the squared differences of predicted and actual data for each predicted day.

Figure(2): Predictive Performance: BMA vs BMOS
Predictive Performance: BMOS vs BMA - Contd…

- From the results, one can see that BMA models the uncertainty in the prediction more accurately than BMOS.
- However, BMA results in greater uncertainty in the prediction when the prediction horizon increases.
- This is reflected in the distributions corresponding to 5 days prediction where box-Whisker plot of BMA covers noticeably wider range than BMOS.
Conclusions

• In this paper, we demonstrated that adaptive MCMC methodologies can be developed and used in CVAR models to estimate parameters and model order.

• Adaptive MCMC methodologies, unlike Griddy-Gibbs samplers, don’t suffer from the curse of dimensionality and thus enable one to design a scalable and automated BCVAR framework to handle large /varying dimension of data.

• We have also demonstrated via real data sets that uncertainties in rank and variation of rank with varying data length might cause serious problems in parameter estimation and prediction.

• Usage of BMA in moment calculations and predictions can remove such uncertainties in ranks and improve the relevant performances.