Model Selection and Adaptive Markov chain Monte Carlo for Bayesian Cointegrated VAR model

15th International Conference on Computing in Economics and Finance.

Gareth W. Peters
CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW, 1670, Australia;
UNSW Mathematics and Statistics Department, Sydney, 2052, Australia;
email: peterga@maths.unsw.edu.au

Balakrishnan Kannan
Baronia Capital Pty. Ltd., 12 Holtermann St., Crows Nest, NSW 2065, Australia.
e-mail: Balakrishnan.Kannan@baroniacapital.com.au

Ben Lasscock
Baronia Capital Pty. Ltd., 12 Holtermann St., Crows Nest, NSW 2065, Australia.
e-mail: ben.lasscock@baroniacapital.com.au

Chris Mellen
Baronia Capital Pty. Ltd., 12 Holtermann St., Crows Nest, NSW 2065, Australia.
e-mail: Chris.Mellen@baroniacapital.com.au

Submitted: 02 June 2009
Abstract

In this paper, we develop novel Markov chain Monte Carlo sampling methodology for Bayesian Cointegrated Vector Auto Regression (CVAR) models. Here we focus on two novel extensions to the sampling methodology for the CVAR posterior distribution. The first extension we develop replaces the popular sampling methodology of the griddy Gibbs sampler with an automated alternative which is based on an Adaptive Metropolis-Hastings algorithm. This is particularly relevant to automate the proposal mechanism in the MCMC algorithm in settings where griddy Gibbs is impractical such as when the dimension of the CVAR series is large, e.g. $d > 5$.

We also treat the rank of the CVAR model as a random variable and perform joint inference on the rank and model parameters. This is achieved with a Bayesian posterior distribution defined over both the rank and the CVAR model parameters, and inference is made via a Savage-Dickey density estimator for the Bayes Factor analysis of rank.

**Keywords:** Cointegrated Vector Auto Regression, Adaptive Markov chain Monte Carlo, Bayesian Inference, Bayes Factors, Savage-Dickey.
1 Introduction

Bayesian analysis of Cointegrated Vector Auto Regression (CVAR) models has been addressed in several papers, see Koop et al. (2006) for an overview of these models. In a CVAR model it is important to carefully specify the priors for posterior model parameters to ensure the posterior is not improper. This has significant implications on the Bayesian model structure. Blind specification of priors on the VAR model coefficients can result in improper posterior distributions, see Koop et al. (2006). For this reason several authors have worked under the Error Correction Model (ECM) framework, see for example p.141-142 of Reinsel and Velu (1998). Hence, there are issues relating to prior specification and then the resultant issue of sampling from the non-standard posterior distribution obtained.

Typically in the Cointegration literature the sampling approach adopted is a Gibbs sampling framework, see Bauwens and Lubrano (1996), Geweke (1996), Kleibergen and van Dijk (1994) and Sugita (2002). In this paper we adopt the model of Sugita (2002) and Geweke (1996). The conjugacy properties of this model result in exact sampling of many of the posterior parameters under a matric-block Gibbs sampler.

This paper extends the matric-variate block Gibbs sampling framework typically used in Bayesian Cointegration models by replacing the computational ”griddy-Gibbs” sampler with two possible automated alternatives which are based on matric-variate adaptive Metropolis-within-Gibbs samplers. Adaptive MCMC is a new methodology to learn on-line the ”optimal” proposal distribution for an MCMC algorithm, see Atachade and Rosenthal (2005), Haario, Saksman and Tamminen (2001; 2007) and Andrieu and Moulines (2006). There are several different versions of adaptive MCMC algorithms. Basically adaptive MCMC algorithms aim to allow the Markov chain to adapt the proposal distribution online throughout the simulation in such a way that the correct stationary distribution is still preserved.

Additionally, we consider rank estimation for reduced rank Cointegration models. From a Bayesian perspective we tackle this via Bayes Factor (BF) analysis for posterior ”model” probabilities of the rank. Then we demonstrate estimation and predictive performance under a Bayesian setting for both Bayesian Model Selection (BMS) and Bayesian Model Averaging (BMA).

The models and algorithms developed allow for estimation of either the rank $r$, i.e. the model index, and the lag $p$ of the CVAR model jointly with the model parameters. Here assume the lag is fixed and known.
In this paper the following notation will be used: ' denotes transpose, \( p(.) \) denotes a density and \( P(.) \) a distribution, \( \Omega \) will be the space on which densities will take their support and it will be assumed throughout that we are working with Lebesgue measure. The operator \( \otimes \) is the Kronecker tensor product which performs element wise matrix multiplication and \( \| \cdot \| \) denotes the total variation norm. We denote generically the state of a Markov chain at time \( j \) by random variable \( \Theta^{(j)} \) and the transition kernel from realized state \( \Theta^{(j-1)} = \theta \) to \( \Theta^{(j)} \) by \( Q(\theta, \Theta^{(j)}) \). In the case of an adaptive transition kernel we will also assume that there is a sequence of transition kernels denoted by \( Q_{\Gamma_j}(\theta, \Theta^{(j)}) \), where \( \Gamma_j \) is the sequence index.

2 CVAR model under ECM framework

We note that a well presented representation to co-integration models is provided by Engle and Granger (1987) and for the original error correction representation of a co-integrated series, see Granger (1981) and Granger (1983). Working with the definitions they provide for an integrated and co-integrated series, we denote \( x_t \) as an integrated of order 1, \( I(1) \), \( n \)-dimensional vector with \( r \) linear cointegrating relationships. The error terms are assumed time independent and zero mean multivariate Gaussian distributed, with covariance \( \Sigma \). In matrix form the model we consider is denoted by,

\[
Y = X\Gamma + Z\beta\alpha' + E = WB + E, \tag{2.1}
\]

where,

\[
\begin{align*}
Y & = \left( \begin{array}{cccc}
\Delta x'_{p} & \Delta x'_{p+1} & \ldots & \Delta x'_{T} \\
\epsilon'_{p-1} & \epsilon'_{p} & \ldots & \epsilon'_{T} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta x'_{T-1} & \ldots & \Delta x'_{T-p+1}
\end{array} \right), \quad Z = \left( \begin{array}{c}
x'_{p-1} \\
x'_{p} \\
\vdots \\
x'_{T-1}
\end{array} \right), \\
E & = \left( \begin{array}{c}
\epsilon'_{p-1} \\
\epsilon'_{p} \\
\vdots \\
\epsilon'_{T}
\end{array} \right), \quad \Gamma = \left( \begin{array}{cccc}
u' & \Psi_{1} & \ldots & \Psi_{p-1}
\end{array} \right), \\
X & = \left( \begin{array}{cccc}
1 & \Delta x'_{p-1} & \ldots & \Delta x'_{1} \\
1 & \Delta x'_{p} & \ldots & \Delta x'_{2} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \Delta x'_{T-1} & \ldots & \Delta x'_{T-p+1}
\end{array} \right), \quad W = \left( \begin{array}{c}
X \ Z\beta
\end{array} \right), \quad B = \left( \begin{array}{c}
\Gamma \ \alpha'
\end{array} \right),
\end{align*}
\]

where the dimensions of these matrices are, \( u \) and \( \epsilon \ (n \times 1) \), \( \Psi \) and \( \Gamma \ (n \times n) \), \( \alpha \) and \( \beta \ (n \times r) \). The parameters \( \mu \) represents the trend coefficients, and \( \Psi_{i} \) is the \( i^{th} \) matrix of autoregressive coefficients and the long run multiplier matrix is given by \( \Pi = \alpha\beta' \). The long run multiplier matrix is an important quantity of this model, its properties include: if \( \Pi \) is a zero matrix, the series \( x_t \) contains \( n \) unit roots; if \( \Pi \) has full rank then the univariate series in \( x_t \) are (trend-)stationary; and co-integration occurs when \( \Pi \) is of rank \( r < n \). The matrix
\( \beta \) contains the co-integration vectors, reflecting the stationary long run relationships between the univariate series in \( x_t \) and the \( \alpha \) matrix contains the adjustment parameters, specifying the speed of adjustment to equilibria \( \beta'x_t \).

This results in a likelihood model, when the parameters of interest are \( B, \Sigma \) and \( \beta \), given by

\[
L(Y|B, \Sigma, \beta) = |\Sigma|^{-t/2} e^{-0.5tr[\Sigma^{-1}(S+R)]},
\]

where \( \Sigma = \text{Cov}(E), t = T-p+1 \) and

\[
R = (B - \hat{B})'W'B(B - \hat{B}), S = (Y - W\hat{B})'(Y - W\hat{B}), \hat{B} = (W'W)^{-1}W'Y
\]

3. Bayesian CVAR models conditional on Rank (r)

The assumptions and restrictions of our Bayesian CVAR model include:

1. **Identification Issue:** For any non-singular matrix \( A \), the matrix of long run multipliers \( \Pi = \alpha\beta' \) is indistinguishable from \( \Pi = \alpha AA^{-1}\beta' \), see Koop, Strachan, Dijk and Villani (2006) or Reinsel and Velu (1998). We use a standard approach to globally overcome this problem by incorporating a non-unique identification constraint. We impose \( r^2 \) restrictions as follows \( \beta = [I_r, \beta_*']' \). However, as noted by Kleibergen and van Dijk (1994) and discussed in Koop, Strachan, van Dijk and Villani (2006) this can still result in local identification issues at the point \( \alpha = 0 \), when \( \beta \) does not enter the model. Hence, one must be careful to ensure that the Markov chain generated by the matrix-variate block Gibbs sampler is not invalidated by the terminal absorbing state. As is standard we monitor the performance of the sampler to ensure this has not occurred.

2. **Error Correction Model:** The ECM framework complicates Bayesian analysis since products, \( \alpha\beta' \), preclude direct use of Monte Carlo samples for inference in the VAR model setting. However, conditional on \( \beta \) the nonlinear ECM model becomes linear and therefore under the informative priors used by Geweke (1995) and Sugita (2002), we can once again apply standard Bayesian analysis to the VAR model.

3. **Prior Choices:** We do not consider the issue of prior distortions illustrated by Kleibergen and van Dijk (1994). This is not the focus of the present paper. Alternative prior models include Jeffrey’s priors, Embedding approach and a focus on the cointegration space.
3.1 Prior and Posterior Model

Here we present the model for estimation of $\beta$, $B$ and $\Sigma$ conditional the rank $r$. As in Sugita (2002), we use a conjugate hierarchical prior.

- $\beta \sim N(\bar{\beta}, Q \otimes H^{-1})$ where $N(\bar{\beta}, Q \otimes H^{-1})$ is the matrix-variate Gaussian distribution with prior mean $\bar{\beta}$, $Q$ is a $(r \times r)$ positive definite matrix, $H$ a $(n \times n)$ matrix.

- $\Sigma \sim IW(S, h)$ where $IW(S, h)$ is the Inverse Wishart distribution with $h$ degrees of freedom and $S$ is an $(n \times n)$ positive definite matrix.

- $B|\Sigma \sim N(P, \Sigma \otimes A^{-1})$ where $N(P, \Sigma \otimes A^{-1})$ is the matrix-variate Gaussian distribution with $h$ degrees of freedom and $S$ is an $(n \times n)$ positive definite matrix.

Combining the priors and likelihood produce matrix variate full conditional posterior distributions: Inverse Wishart distribution for $\Sigma|\beta, Y$; matrix-variate student-t distribution for $B|\beta, Y$ (or alternatively a matrix-variate Gaussian for $B|\beta, \Sigma, Y$); and for $\beta|Y$,

$$p(\beta|Y) \propto p(\beta)|S_x|^{-(t+h+1)/2}|A_s|^{-n/2}.$$ \hspace{1cm} (3.1)

See Sugita (2002) for derivations.

3.2 Sampling and Estimation Conditional on Rank $r$

Here we focus on obtaining samples from the posterior distribution which can be used to obtain Bayesian parameter estimates (MMSE, MAP). The complication in sampling arises with the full conditional posterior 3.1 which can not be sampled from via straight forward inversion sampling.

In this paper we outline novel algorithms to sample from the posterior distribution $p(\beta|Y, B, \Sigma, r)$, providing an alternative automated approach to the griddy-Gibbs sampler algorithm made popular in this Bayesian co-integration setting by Bauwens and Lubrano (1996).

The matrix-variate griddy-Gibbs sampler numerically approximates the target posterior on a grid of values and then performs numerical inversion to obtain samples from 3.1 at each stage of the MCMC algorithm. Such a grid based procedure will suffer from the curse of dimensionality when $n$ is large ($n > 5$) after which it becomes highly inefficient. Note additionally, alternatives such as Importance Sampling will also be problematic once $n$ becomes too large. It is difficult to optimize the choice of the Importance Sampling distribution which will minimize the variance in the importance weights.

We propose alternative samplers using adaptive matrix-variate MCMC methodology. They do not suffer from the curse of dimensionality and are simple to implement and automate.
• **Algorithm 1 - Random Walk (mixture local & global moves):** Involves an offline adaptively pretuned mixture proposal containing a combination of local and global Random Walk (RW) moves. The proposal for the local RW moves have standard deviation tuned to produce average acceptance probabilities between [0.3, 0.5]. The independent global matrix-variate proposal updates all elements of $\beta$ via a multivariate Gaussian proposal centered on Maximum Likelihood parameter estimates for $\beta$ and the Fisher information matrix for the covariance of the global proposal. This is similar to the approach adopted in Vermaak et al. (2004) and Fan et al. (2008).

• **Algorithm 2 - Adaptive Random Walk:** Involves an online matrix-variate adaptive Metropolis algorithm based on methodology presented in Roberts and Rosenthal (2008).

Proceeding sections denote the algorithmic 'time' index by $j$ and the current state of a Markov chain for generic parameter $\theta$ at time $j$ by $\theta^{(j)}$. The length of the Markov chain is $J$.

Note, since we have imposed $r^2$ restrictions in the form of $I_{r \times r}$, any proposal for $\beta = [I, \tilde{\beta}]$ will only correspond to the unrestricted elements of $\beta$ denoted by $\tilde{\beta}$. In our case, these correspond to those in locations $(n - r) \times r$.

### 3.2.1 Algorithm 1

In Algorithm 1 the mixture proposal distribution for parameters $\tilde{\beta}$ will be given by,

$$q\left(\tilde{\beta}^{(t-1)}, \cdot \right) = w_1 N \left( \tilde{\beta}; \tilde{\beta}_{\text{ML}}, \Sigma_{\text{ML}} \right) + (1 - w_1) N \left( \tilde{\beta}_{i,k}; \tilde{\beta}_{i,k}^{(t-1)}, \sigma_{i,k}^2 \right). \quad (3.2)$$

The Maximum Likelihood parameters are obtained off-line, see (p. 286 Lutkepohl (2007)). The local random walk proposal variances $\sigma_{i,k}^2$ for each element of $\tilde{\beta}$ are obtained via pre-tuning.

### 3.2.2 Algorithm 2: Adaptive Metropolis within Gibbs sampler moves for CVAR model given rank $r$

There are several classes of adaptive MCMC algorithms, see Roberts and Rosenthal (2008). The distinguishing feature of adaptive MCMC algorithms, compared to standard MCMC, is generation of the Markov chain via a sequence of transition kernels. Adaptive algorithms utilize a combination of time or state inhomogeneous proposal kernels. Each proposal in the sequence is allowed to depend on the past history of the Markov chain generated, resulting in many variants.

Due to the inhomogeneity of the Markov kernel used in adaptive algorithms, it is particularly important to ensure the generated Markov chain is ergodic, with the appropriate stationary
Algorithm 1: MH within Gibbs sampler for fixed rank $r$ via a pretuned mixture of global and local moves.

**Input:** Initial Markov chain state $(\Sigma(0), B(0), \beta(0))$.

**Output:** Markov chain samples $\{\Sigma(j), B(j), \beta(j)\}_{j=1:J} \sim p(\Sigma, B, \beta|Y)$.

**begin**

1. Set initial state $(\Sigma(0), B(0), \beta(0))$ deterministically or by sampling the priors.

2. Calculate Maximum Likelihood parameters $\tilde{\beta}^{ML}$ and $\Sigma^{ML}$.

3. Initialize $w_1$ and $w_2 = 1 - w_1$ and index $j = 1$.

**repeat**

5. Sample $\Sigma$ via inversion to obtain $\Sigma(j)$.

6. Sample $B$ via inversion to obtain $B(j)$.

7. Sample realization $U = u$ where $U \sim U[0, 1]$

   **if** $u \geq w_1$ **then** /* perform a local random walk move */

   7a. Sample uniformly index $(i, k)$ from set of $n - r \times r$ elements.

   7b. Sample the $(i, k)$-th component $\tilde{\beta}_{i,k}^* \sim N(\tilde{\beta}_{i,k}; \tilde{\beta}_{i,k}^{(j-1)}, \sigma_{i,k}^2)$.

   7c. Construct proposal $\beta^* = [I_{r \times r}, \tilde{\beta}^*]$, where $\tilde{\beta}^*$ is $\tilde{\beta}^{(j-1)}$ with the $(i, k)$-th element given by $\tilde{\beta}_{i,k}^*$.

   **else** /* perform a global independent move */

   7a. Sample proposal $\tilde{\beta}^* \sim N(\tilde{\beta}; \tilde{\beta}^{ML}, \Sigma^{ML})$.

   7b. Construct proposal $\beta^* = [I_{r \times r}, \tilde{\beta}^*]$.

8. Calculate Metropolis Hastings Acceptance Probability:

   $$A\left(\beta^{(j-1)}, \beta^*\right) = \frac{p\left(\Sigma(j), B(j), \beta^*|Y\right) q\left(\beta^* \rightarrow \beta^{(t-1)}\right)}{p\left(\Sigma(j), B(j), \beta^{(j-1)}|Y\right) q\left(\beta^{(t-1)} \rightarrow \beta^*\right)}$$ (3.3)

   Accept $\beta^{(j)} = \beta^*$ via rejection using $A$, otherwise $\beta^{(j)} = \beta^{(j-1)}$.

9. $j = j + 1$

**until** $j = J$

**end**
distribution. Several recent papers proposing theoretical conditions that must be satisfied to ensure ergodicity of adaptive algorithms include, Atchade and Rosenthal (2005), Roberts and Rosenthal (2008), Haario et al. (2007), Andrieu and Moulines (2006) and Andrieu and Atchade (2007).

Haario et al. (2001) developed an adaptive Metropolis algorithm with proposal covariance adapted to the history of the Markov chain. The original proof of ergodicity of the Markov chain under such an adaption was overly restrictive. It required a bounded state space and a uniformly ergodic Markov chain.

Roberts and Rosenthal (2005) proved ergodicity of adaptive MCMC under simpler conditions known as Diminishing Adaptation and Bounded Convergence. As in Roberts and Rosenthal (2005) we assume that each fixed kernel in the sequence \( Q \) has stationary distribution \( P(\cdot) \).

Define the convergence time for kernel \( Q_j \) when starting from state \( \theta \) as \( M_{\epsilon}(\theta, \gamma) = \inf\{j \geq 1 : \|Q^j_\gamma(\theta, \cdot) - P(\cdot)\| \leq \epsilon \} \). Under these assumptions, they derive the sufficient conditions;

- **Diminishing Adaptation:** \( \lim_{j \to \infty} \sup_{\theta \in E} \| Q_{j+1}(\theta, \cdot) - Q_j(\theta, \cdot) \| = 0 \) in probability.
  
  Note, \( \Gamma_j \) are random indices.

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

which guarantee asymptotic convergence in two senses,

- Asymptotic convergence: \( \lim_{j \to \infty} \| L(\Theta(j)) - P(\cdot) \| = 0 \)

- WLLN: \( \lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^j g(\Theta(i)) = \int g(\theta)p(\theta)d\theta \) for all bounded \( g : E \to \mathbb{R} \).

It is non-trivial to develop adaption schemes which can be verified to satisfy these two conditions. We develop a matric-variate adaptive MCMC methodology in the CVAR setting, using a proposal kernel known to satisfy these two ergodicity conditions for unbounded state spaces and general classes of target posterior distribution, see Roberts and Rosenthal (2008) for details.

In Algorithm 2 the mixture proposal distribution for parameters \( \tilde{\beta} \) which is \( d = (n - r) \times r \) dimensional and is given at iteration \( j \) by,

\[
q_j\left(\tilde{\beta}_{(t-1)}, \cdot \right) = w_1 N\left(\tilde{\beta}_t; \tilde{\beta}_{(t-1)}, \frac{(2.38)^2}{d} \Sigma_j\right) + (1 - w_1) N\left(\tilde{\beta}_t; \tilde{\beta}_{(t-1)}, \frac{(0.1)^2}{d} I_{d,d}\right),
\]

(3.4)

Here, \( \Sigma_j \) is the current empirical estimate of the covariance between the parameters of \( \tilde{\beta} \) estimated using samples from the Markov chain up to time \( j \). The theoretical motivation for the
choices of scale factors 2.38, 0.1 and dimension d are all provided in Roberts and Rosenthal (2008) and are based on optimality conditions presented in Roberts et al. (1997) and Roberts and Rosenthal (2001). The adaptive MCMC Algorithm 2 is identical to Algorithm 1 except we replace step 7 with the following alternative;

Algorithm 2: matric-variate adaptive MH within Gibbs sampler for fixed rank r.

```plaintext
if \( u \geq w_1 \) then /* perform an adaptive random walk move */
    7a. Estimate \( \Sigma_j \) the empirical covariance of \( \beta \) for elements in \((n - r) \times r\) using samples \( \{\tilde{\beta}^{(i)}\}_{i=1}^{j} \).
    7b. Sample proposal \( \tilde{\beta}^* \sim N \left( \tilde{\beta}; \tilde{\beta}^{(t-1)}, \frac{(2.38)^2}{d}\Sigma_j \right) \).
    7c. Construct proposal \( \beta^* = [I_{r \times r}, \tilde{\beta}^*] \).
else /* perform a non-adaptive random walk move */
    7a. Sample proposal \( \tilde{\beta}^* \sim N \left( \tilde{\beta}; \tilde{\beta}^{(t-1)}, \frac{(0.1)^2}{d}I_{d,d} \right) \).
    7b. Construct proposal \( \beta^* = [I_{r \times r}, \tilde{\beta}^*] \).
```

4 Rank Estimation for Bayesian VAR Cointegration models

Here we discuss the Bayes Factor approach to rank estimation which is inefficient since it involves running \( d+1 \) Markov chains, one for each model (rank \( r \)). For a sophisticated alternative which presents a novel TD-MCMC based approach, requiring a single Markov chain to obtain samples from the posterior distribution \( p(B, \Sigma, \beta, r|Y) \), see Peters et al. (2009).

4.1 Posterior Model Probabilities for Rank \( r \) via Bayes Factors

In Sugita (2002) and Kleibergen and Paap (2002) the rank is estimated via Bayes factors, a popular approach to Bayesian model selection in Bayesian cointegration literature. We note that alternative approaches to rank estimation include Strachan and van Dijk (2004). Sugita (2002) works with a conjugate prior on \( \alpha \) which will not produce a problem with Bartlett’s paradox, posterior probabilities of the rank are well defined.
4.1.1 Bayes Factors via Savage-Dickey density estimator

Sugita (2002) compares the rank of the unrestricted $\alpha$ to the 0 rank setting. Note, Kleibergen and Pap (2002) have a slightly different approach in that they compared each rank $r$ to the full rank case for the unrestricted $\alpha$ parameter.

Under a rank 0 comparison the posterior model probabilities are given by,

$$P_r (r|Y) = \frac{BF_{r|0}}{\sum_{j=0}^{n} BF_{j|0}},$$

(4.1)

with $BF_{0|0}$ defined as 1.

In the calculation of $BF_{r|0}$, Sugita (2002) recommends a generalized Savage-Dickey density ratio, first introduced by Verdinelli and Wasserman (1995) for nested model structure Bayes factors.

$$BF_{r|0} = \frac{p(\alpha' = 0_{r \times n})}{C_{r}^{-1} p(\alpha' = 0_{r \times n}|Y)} = \frac{\int p(\alpha, \beta, \Gamma, \Sigma|Y)d\alpha d\beta d\Gamma d\Sigma}{C_{r}^{-1} \int p(\alpha, \beta, \Gamma, \Sigma|rank(\alpha) = 0|Y)d\alpha d\beta d\Gamma d\Sigma}$$

(4.2)

where the correction factor for the reduction in dimension $C_{r}$ is given by,

$$C_{r} = \int p(\alpha, \beta, \Gamma, \Sigma)|_{rank(\alpha) = 0}d\beta d\Gamma d\Sigma.$$

(4.3)

We note that Sugita (2002) does not comment on numerical complications that can arise when implementing the Savage Dickey estimator for the CVAR model. We detail in Appendix 1, Section 7 steps that were critical to the calculation of the Bayes Factors when handling potential numerical overflows. The numerical issues arise as $t$ increases, for example the term $|S_{t}^{(i)}|^{1+h}$ will explode numerically. This will result in incorrect numerical results for the Bayes Factors if not handled appropriately.

4.2 Model Selection, Model Averaging and Prediction

With samples from $p(\beta, B, \Sigma, r|Y)$ one can consider either model selection or model averaging. In a survey of the literature on rank selection, the most common form of inference performed involves model selection. In this paper we note that model averaging should also be considered, especially when it is probable that given the realized data, two different ranks are highly probable according to their posterior model probabilities. In this case one can use the samples from $p(\beta, B, \Sigma|Y, r)$ in each model $r$ to form a weighted model averaged estimate through the direct knowledge of the estimated model probabilities given by $p(r|Y)$. There is discussion on model averaging in the CVAR context found in Koop, Strachan, van Dijk and Villani (2006).
Bayesian Model Order Selection (BMOS)
In BMS we select the most probable model corresponding to the maximum \textit{a posteriori} (MAP) estimate from $p(r|Y)$, denoted $r_{MAP}$. Conditional on $r_{MAP}$, we then take the samples of $\{\beta^{(j)}, B^{(j)}, \Sigma^{(j)}\}_{j=1:M}$ corresponding to when the transdimensional Markov chain visited model $r_{MAP}$ and we estimate point estimates for the parameters.

These point estimates typically include posterior means or modes, though one should be careful. We note that it was demonstrated by Kleibergen and van Dijk (1994) or Bauwens and Lubrano (1996) that in many popular CVAR Bayesian models, certain choices of prior result in a proper posterior yet it may not have finite moments of any order. Some alternatives are proposed by Strachan and Inder (2004).

Bayesian Model Averaging (BMA)
In this section we consider the problem of estimating for example an integral of a quantity or function of interest, $\phi(\{\beta, B, \Sigma\})$, with respect to the posterior distribution of the parameters, e.g. moments of the posterior. Since we have chosen to work with a posterior distribution $p(\beta, B, \Sigma, r|Y)$ we can estimate this integral quantity whilst removing the model risk associated with rank uncertainty. This is achieved by approximating

$$
\sum_{r=1}^{n} \int \phi(\{\beta, B, \Sigma|r\}) p(\beta, B, \Sigma|Y, r)p(r|Y)d\beta dBd\Sigma \approx \sum_{r=1}^{n} \sum_{j=1}^{M} \phi(\{\beta^{j}, B^{j}, \Sigma^{j}|r^{j}\}) p(r^{j}|Y). \quad (4.4)
$$

Prediction Incorporating Model Risk
Here we perform prediction whilst removing model uncertainty related to the rank. This is possible under a Bayesian Model Averaging (BMA) framework using,

$$
p(Y^*|Y) = \sum_{r} \int p(Y^*|\beta, B, \Sigma, r)p(\beta, B, \Sigma|Y, r)p(r|Y)d\beta dBd\Sigma. \quad (4.5)
$$

We will compare the predictive performance of the MMSE estimate or mean of the estimated distribution for $p(Y^*|Y)$ under the BMA versus BMOS approach which involves,

$$
p(Y^*|Y) = \int p(Y^*|\beta, B, \Sigma, \hat{r}_{MAP}) p(\beta, B, \Sigma|Y, \hat{r}_{MAP}) d\beta dBd\Sigma. \quad (4.6)
$$

5 Simulation Experiments
Analysis of the methodology developed is in three parts: the first part contains simulations performed on synthetic data sets, comparing performance of the proposed model sampling method-
ogy; the second part contains two real data set examples; and the third part involves analysis of predictive performance BMOS and BMA using real data.

5.1 Synthetic Experiments

In this section the intention will be to develop a controlled setting in which the true model parameters are known and the data is generated from the true model. This will allow us to assess performance of each of the proposed estimation procedures. In doing this we take an identical model to the toy model studied in Sugita (2002) for our analysis.

5.1.1 Analysis of samplers

The first analysis is to compare the performance of the two adaptive samplers. To achieve this we generate 20 realizations of data sets of length $T = 100$ from the rank $r = 2$ model. Then conditional on knowledge of the rank $r = 2$ we sample $J = 20,000$ samples from the joint posterior $p(B, \Sigma, \beta | Y, r = 2)$ and discard the first 10,000 samples as burnin. We perform this analysis for each of the data realizations under both of the proposed samplers, Algorithm 1 and Algorithm 2, and then we present average MMSE estimates and average posterior standard deviations from each sampler in Table 1. In particular we present the averaged posterior point estimates for: the unrestricted $\beta$ parameters; the average trace of the posterior estimate of the covariance $\Sigma$; the average of each of the intercept terms; and the averaged first element of the unrestricted $\alpha$.

Note, the pre-tuning of the local random walk proposal standard deviation for Algorithm 1 is performed offline using an MCMC run of length 20,000. Additionally, the prior parameters were set to be: for $B | \Sigma$ the prior parameters were set as $P = (\hat{W}^T\hat{W})^{-1}\hat{W}Y$, $A = \lambda (\hat{W}^T\hat{W})/T$ with $\lambda = 1$, $\hat{W} = (XZ\hat{\beta})$ and $\hat{\beta} = [I_{r \times r}, 0]$; for $\beta$ the prior parameters were set as $E[\beta] = (I_{r \times r}, 0)$, $Q = I_{n \times n}$, $H = \tau Z'Z$ and $\tau = 1/T$; for $\Sigma$ the prior parameters were set as $S = \tau Y'Y$ and $h = n + 1$.

These results demonstrate that both Algorithm 1 and Algorithm 2 perform well. The MMSE estimates produced by both algorithms are accurate compared to the true parameter values used to generate the data. Algorithm 1 which involved the mixture of pretuned local moves and a Global move centered on the Maximum Likelihood parameter estimates required more computational effort than the adaptive MCMC approach of Algorithm 2. Additionally, we point out that as discussed in Rosenthal (2008), the sampler we developed in Algorithm 2 actually achieves optimal performance as $d \to \infty$. Therefore it will be a far superior algorithm to the
griddy Gibbs sampler approach which will not be feasible in high dimensions. Hence, for an automated and computationally efficient alternative to the Griddy Gibbs sampler typically used we would recommend the use of Algorithm 2. In the following studies, we utilize Algorithm 2, the adaptive MCMC algorithm.

5.1.2 Analysis of model selection in the Bayesian CVAR model

In this section we study on synthetic data the performance of Savage-Dickey density estimator applied to estimate posterior model probabilities for the rank via the Bayes Factor methodology presented earlier. To perform this analysis we take data series of length $T = 100$ and we simulate 50 independent data realizations for each possible model rank $r = 1, \ldots, 4$. Then for each rank $r$ we count the number of times each model is selected as the MAP estimate out of the total of the 50 simulations, one simulation per generated data set. Note, the algorithm was run for 20,000 iterations with 10,000 samples used as burnin. The results of this analysis are presented in Table 2.

We note that their results of this section demonstrated the following interesting properties:

1. When the true rank used to generate the observations data was small, the BF methodology was clearly able to detect the true model order as the MAP estimate in a high proportion of the tested data sets.

2. In all cases the averaged actual posterior model probabilities were very selective of the correct model, indicating that at least under this synthetic data scenario, there would not be great benefit in performing model averaging. However, we will demonstrate later examples with actual data in which there is significant ambiguity between possible model ranks, in these cases we also study the model averaging results.

5.2 Financial Example 1 - US mini indexes

Having assessed the proposed algorithms developed in this paper for synthetic data generated from a CVAR model, we now work with a practical financial example. In this example we will consider data series comprised of US indexes S&P mini, Nasdaq mini and Dow Jones mini. The data obtained for each of these data series consists of 774 values corresponding to the close of market daily price from the 31-Aug-2005 through to 30-Sep-2008. The time series data is presented in Figure 7.
We analyze this data using Algorithm 2 (adaptive MCMC) and estimate the rank via Bayes Factor analysis. The results are presented in Table 3, where the rank is estimated via Bayes-Factors. We run 20 independent samplers with different initializations, for each possible rank. This is performed for each data set, and the total series is split into increasing subsets, each taking subsets of the data from 50 data points through to 400 data points, in increases of 50 data points. This allows us to study the change in the estimated rank as a function of time for each of these time series. Clearly, if the true rank of our model was fixed, then as the total amount of data we include increases, then we should see the posterior model probability of the rank converge to 1 for one of the possible ranks. What we observed after doing this analysis was that there was a clear variability in the predicted rank as we included more data. In particular the model gave preference most often to rank 1, suggesting that 2 common trends are present in the series. Additionally, the fact that in several cases, the model is less likely to distinguish between rank 1 and 2, suggests it may be prudent to also perform a model averaging analysis.

5.3 Financial Example 2 - US notes

Here we repeat the same procedure performed in Financial Example 1, for a different data set. This time we consider data series comprised of Bond data for US 5 year, 10 year and 30 year notes over the same time period as the US mini index data. The time series data is presented in Figure 7.

We analyze this data using Algorithm 2 (adaptive MCMC) and estimate the rank via Bayes Factor analysis. The results are presented in Table 4, where the rank is estimated via Bayes-Factors. We run 20 independent samplers with different initializations, for each possible rank. This is performed for each data set, and the total series is split into increasing subsets, each taking subsets of the data from 50 data points through to 400 data points, in increases of 50 data points. This allows us to study the change in the estimated rank as a function of time for each of these time series. Again, we observed that with this data, the model gave preference most often to rank 1, suggesting that 2 common trends are present in the series we area analysis. However, there was much stronger evidence for a single co-integrating relationship over time in this data, compared to the analysis of the US mini index data over the same period. This suggests that the US bond data series is a more stable series to fit the CVAR model too when assuming a constant number of co-integrating relationships over time.
5.4 Financial Example 3

In this section we perform a predictive performance comparison using Bayesian Model Selection versus Averaging. We take 2 series for the US bonds, 5 years and 10 years, and we combine these series over the same period with the S&P 500 mini index. We compare the MMSE estimate of the predicted series over 10 steps ahead which is obtained from the distribution of the predicted data $p(Y^*|Y)$, after we have integrated out parameter and rank uncertainties. We demonstrate that in this actual data example, the performance obtained by Bayesian Model Averaging represents the uncertainty in the prediction more accurately than the Bayesian Model Order Selection setting.

This study is performed as follows. We begin by selecting randomly, with replacement, 100 segments of the vector time series, each containing 50 days of data. For each segment of the time series we fit our Bayesian model for each possible rank, also estimating via Bayes Factors the posterior model probability for each rank. Then we calculate the predictive posterior mean, corresponding to the MMSE estimate of the predicted data series for the following 5 days, $Y^*$. Finally, we take the squared difference between the actual data series over the proceeding 10 days post the 50 days for the given segment and the posterior mean of the predicted data $Y^*$.

In Figure 7 we present for each prediction day a histogram of the squared difference between the actual data over the random sets of 5 days and the predictive posterior MMSE estimators for the same 5 days. We compare here the performance under Bayesian model selection and averaging. When performing Bayesian model averaging we are integrating out uncertainty in the prediction due to the prediction of the unknown rank.

Clearly, the Bayesian model averaging approach will result in a greater uncertainty in the prediction when compared to the Bayesian model selection. This is reflected especially in the distribution of the prediction at 5 days where the model averaging approach box-whisker plot covers a noticeably wider range than the model selection equivalent. Though not presented here, we also assessed and confirmed this would occur out to longer predictions of 10 days and 20 days.

6 References

1. Ackert, L., Racine, M.D. (1998). "Stochastic trends and cointegration in the market for equities". Federal Reserve Bank of Atlanta, Working Paper 98-13.

2. Andrieu, C., Moulines, E. (2006). "On the ergodicity properties of some adaptive MCMC algo-
3. Andrieu, C., Atchade, Y. (2007). "On the efficiency of adaptive MCMC algorithms". *Electronic Communications in Probability*, 12, 336-349.

4. Atchade, Y.F., Rosenthal, J.S. (2005). "On adaptive Markov chain Monte Carlo algorithms". *Bernoulli*, 11(5), 815-828.

5. Bauwens, L., Lubrano, M., Richard, J.F. (1999). "Bayesian Inference in Dynamic Econometric Models: Advanced Texts in Econometrics". Oxford University Press.

6. Bauwens, L., Lubrano, M. (1996). "Identification restrictions and posterior densities in cointegrated Gaussian VAR systems". *Advances in Econometrics* 11, Part B, (JAI Press, Greenwich) 3-28.

7. Bauwens, L., Giot, P. (1997). "A Gibbs sampling approach to cointegration". *Computational Statistics*, 13, 339-368.

8. Engle, R.F., Granger, C.W.J. (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing". *Econometrica*, 55(2), 251-276.

9. Geweke J. (1996). "Bayesian reduced rank regression in econometrics". *Journal of Econometrics*, 75, 121-146.

10. Granger, C.W.J. (1981). "Some properties of time series data and their use in econometric model specification". *Journal of Econometrics*, 121-130.

11. Granger, C.W.J., Weiss, A.A. (1983). "Time series analysis of error-correcting models". *Studies in econometrics, time series, and multivariate statistics*. New York: Academic Press, 255-278.

12. Haario, H., Saksman, E., Tamminen, J. (2007). "Componentwise adaptation for high dimensional MCMC". *Computational Statistics*, 20(2).

13. Haario, H., Saksman, E., Tamminen, J. (2001). "An adaptive metropolis algorithm". *Bernoulli*, 7, 223-242.

14. Kleibergen, F., Paap, R. (2002). "Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration". *Journal of Econometrics*, Elsevier, 111(2), 223-249.

15. Kleibergen, F., van Dijk, H.K. (1994). "On the Shape of the Likelihood/Posterior in Cointegration Models". *Econometric Theory*, Cambridge University Press, 10(3-4), 514-551.

16. Koop, G., Strachan, R., van Dijk, H., Villani, M. (2006). "Bayesian Approaches to Cointegration". In T.C.Mills and K. Patterson (Ed.), *Palgrave Handbook of Econometrics Volume 1 Econometric Theory 1st ed.* (pp. 871-898) UK: Palgrave Macmillan.

17. Krolzig, H. M. (1996). "Statistical analysis of cointegrated VAR processes with Markovian regime shifts". *Institute of Economics and Statistics, Oxford Technical Report.*
18. Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer

19. Peters, G.W., Kannan, B.K., Lasscock, B., Mellen, C. (2009). "Trans-dimensional and adaptive Markov chain Monte Carlo for Bayesian co-integrated VAR models". Technical Report: University of NSW, Statistics Department.

20. Reinsel, G.C., Velu, R.P. (1998). *Multivariate Reduced-Rank Regression, Theory and Applications*. Lecture Notes in Statistics Springer.

21. Roberts, G.O., Rosenthal, J.S. (2006). "Examples of Adaptive MCMC". *Journal of Computational and Graphical Statistics*, to appear.

22. Roberts, G.O., Gelman, A., Gilks, W.R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms". *Annals of Applied Probability*, 7, 110-120.

23. Roberts, G.O., Rosenthal, J.S. (2001). "Optimal scaling for various Metropolis-Hastings algorithms". *Statistical Science*, 16, 351-367.

24. Rosenthal, J.S. (2008). "Optimal Proposal Distributions and Adaptive MCMC". Chapter for MCMC Handbook, Brooks S., Gelman A., Jones G. and Meng X.L. eds.

25. Strachan, R.W., van Dijk, H.K. (2004). "Bayesian model selection with an uninformative prior". *Keele Economics Research Papers*

26. Strachan, R.W., Inder, B. (2004). "Bayesian analysis of the error correction model". *Journal of Econometrics*, 123(2), 307-325.

27. Sugita, K. (2002). "Testing for cointegration rank using Bayes factors". *Royal Economic Society Annual Conference*, Royal Economic Society.

28. Villani, M. (2005). "Bayesian reference analysis of cointegration". *Econometric Theory*, (21) 326-357.
We begin by calculating the log posterior model probabilities,

$$\log (Pr (r|Y)) = \log (BF_{r|0}) + \log (BF_{max|0}) - \log \left( \sum_{j=0}^{n} \exp \left( \log (BF_{j|0}) - \log (BF_{max|0}) \right) \right),$$

(7.1)

where $BF_{max|0} = \max \{BF_{0|0}, ..., BF_{n|0}\}$. Additionally, we now consider the log of the Bayes Factor for rank $r$ and we apply the same numerical trick.

$$\log (BF_{r|0}) = \log (p(\alpha = 0_{r \times n})) + \log (C_r) - \log (p(\alpha = 0_{r \times n}|Y))$$

(7.2)

Now, considering each of the terms:

- $\log (p(\alpha = 0_{r \times n})) = -\frac{nr}{2} \log (\pi) + \frac{h}{2} \log (|S|) + \frac{n}{2} \log (|A_{22.1}|) + \sum_{j=1}^{n} \log \left( \frac{\Gamma\left(\frac{h+r+1-j}{2}\right)}{\Gamma\left(\frac{h+1-j}{2}\right)} \right) - \frac{h+r}{2} \log (|S|)$

- $\log (p(\alpha = 0_{r \times n}|Y)) = -\log (N) + \log \left( L_{max}^{(1)} \right) - \log \left( \sum_{i=1}^{N} \log \left( L_i^{(1)} \right) - \log \left( L_{max}^{(1)} \right) \right)$, where $L_i^{(1)} = \pi^{-\frac{nr}{2}} |S^{(i)}|^{-\frac{h+r}{2}} |A_{22.1}^{(i)}|^{-\frac{n}{2}} \prod_{i=1}^{n} \Gamma\left(\frac{h+r+1-i}{2}\right) |S^{(i)} + B_{22}^{(i)} A_{22.1}^{(i)} B_{22}^{(i)}|^{-\frac{h+r}{2}}$ and $L_{max}^{(1)} = \max \{L_1^{(1)}, ..., L_N^{(1)}\}$.

- $\log (C_r) = -\log (N) + \log \left( L_{max}^{(2)} \right) \log \left( \sum_{i=1}^{N} \log \left( L_i^{(2)} \right) - \log \left( L_{max}^{(2)} \right) \right)$, where $L_i^{(2)} = \frac{p(\alpha = 0, \Gamma^{(i)}|\Sigma^{(i)})}{p(\Gamma^{(i)}|\Sigma^{(i)})}$ and $L_{max}^{(2)} = \max \{L_1^{(2)}, ..., L_N^{(2)}\}$. Note this sum evaluated using samples from the Markov chain run in model $r$ where, $p(\alpha = 0, \Gamma^{(i)}|\Sigma^{(i)})$ and $p(\Gamma^{(i)}|\Sigma^{(i)})$ are obtained using knowledge of the specified prior, $p(B|\Sigma) = p(\Gamma, \alpha|\Sigma) = p(\mu, \Psi_{1:p-1}, \alpha|\Sigma)$. 

19
| Parameter Estimates               | Algorithm 1  | Algorithm 2  | Truth |
|----------------------------------|--------------|--------------|-------|
| Ave. MMSE $\beta_{1,r+1}$       | -0.002 (0.001) | -0.034 (0.002) | 0     |
| Ave. Posterior Stdev. $\beta_{1,r+1}$ | 0.018 (0.006) | 0.010 (0.003) | -     |
| Ave. MMSE $\beta_{2,r+1}$       | -0.819 (0.051) | -0.862 (0.045) | -1    |
| Ave. Posterior Stdev. $\beta_{2,r+1}$ | 0.032 (0.005) | 0.020 (0.003) | -     |
| Ave. MMSE $\beta_{1,n}$         | 0.033 (0.025)  | -0.024 (0.023) | 0     |
| Ave. Posterior Stdev. $\beta_{1,n}$ | 0.030 (0.012) | 0.026 (0.010) | -     |
| Ave. MMSE $\beta_{2,n}$         | -0.752 (0.098) | -0.774 (0.082) | -1    |
| Ave. Posterior Stdev. $\beta_{2,n}$ | 0.038 (0.013) | 0.028 (0.006) | -     |
| Ave. Mean acceptance probability $\beta$ | 0.352 (0.010) | 0.232 (0.029) | -     |
| Ave. MMSE tr ($\Sigma$)         | 4.945 (0.331)  | 4.432 (0.332) | 4     |
| Ave. Posterior Stdev. tr ($\Sigma$) | 0.420 (0.049) | 0.416 (0.048) | -     |
| Ave. MMSE $\mu_1$               | 0.07 (0.051)   | 0.065 (0.043) | 0.1   |
| Ave. Posterior Stdev. $\mu_1$   | 0.236 (0.028)  | 0.226 (0.026) | -     |
| Ave. MMSE $\mu_2$               | -0.027 (0.041) | -0.034 (0.024) | 0.1   |
| Ave. Posterior Stdev. $\mu_2$   | 0.183 (0.041)  | 0.181 (0.010) | -     |
| Ave. MMSE $\mu_3$               | -0.080 (0.084) | -0.061 (0.045) | 0.1   |
| Ave. Posterior Stdev. $\mu_3$   | 0.199 (0.020)  | 0.187 (0.015) | -     |
| Ave. MMSE $\mu_4$               | 0.024 (0.049)  | 0.030 (0.029) | 0.1   |
| Ave. Posterior Stdev. $\mu_4$   | 0.184 (0.010)  | 0.185 (0.011) | -     |
| Ave. MMSE $\alpha_{1,1}$        | -0.223 (0.015) | -0.224 (0.016) | -0.2  |
| Ave. Posterior Stdev. $\alpha_{1,1}$ | 0.070 (0.006) | 0.068 (0.005) | -     |
| Ave. MMSE $\alpha_{1,2}$        | 0.201 (0.013)  | 0.202 (0.013) | 0.2   |
| Ave. Posterior Stdev. $\alpha_{1,2}$ | 0.053 (0.002) | 0.052 (0.002) | -     |

Table 1: **Sampler Analysis** - Algorithm 1 is the pretuned mixture proposal of Global ML move and local pretuned MCMC move; Algorithm 2 is the Global adaptively learnt MCMC proposal. Averages and a standard error are taken for the Bayesian point estimators over 20 data sets, the standard errors are presented in brackets (·). Note in all simulations the initial Markov chain is started very far away from the true parameter values.
| Model Rank | Bayes Factors |
|------------|---------------|
| $r = 0$    | 3 (0.84)      |
| $r = 1$    | 16 (0.93)     |
| $r = 2$    | 2 (0.92)      |
| $r = 3$    | 0 (-)         |
| $r = 4$    | 0 (-)         |
| $r = 0$    | 0 (-)         |
| $r = 1$    | 5 (0.89)      |
| $r = 2$    | 13 (0.91)     |
| $r = 3$    | 0 (-)         |
| $r = 4$    | 2 (0.92)      |
| $r = 0$    | 0 (-)         |
| $r = 1$    | 0 (-)         |
| $r = 2$    | 4 (0.89)      |
| $r = 3$    | 6 (0.90)      |
| $r = 4$    | 10 (0.94)     |
| $r = 0$    | 0 (-)         |
| $r = 1$    | 0 (-)         |
| $r = 2$    | 0 (-)         |
| $r = 3$    | 2 (0.87)      |
| $r = 4$    | 18 (0.89)     |

Table 2: **Between Model Analysis** - The true model rank used to generate the data is presented in bold. TDMCMC is the Trans-dimensional Markov chain Monte Carlo algorithm utilizing adaptive MH within model moves and the global Independent between model moves. The results represent the total number of times a given rank is selected as the MAP estimate out of the 20 independent data sets, each of length $T=100$, analyzed. Additionally, the average posterior model probability for these cases is presented in brackets.
| Rank \ T | 50     | 100    | 150    | 200    | 250    | 300    | 350    | 400    |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| r = 0   | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| r = 1   | 8.09 (0.78) | 3.77 (0.29) | 7.01 (0.50) | 11.51 (0.90) | 1.69 (0.54) | 2.71 (3.55) | 3.14 (1.11) | 7.77 (0.83) |
| r = 2   | 2.91 (1.24) | 2.33 (1.26) | 4.61 (0.63) | 25.36 (7.19) | -5.33 (1.17) | -5.80 (0.97) | 4.92 (1.06) | -3.88 (1.11) |
| r = 3   | -26.03 (1.06) | -8.45 (0.27) | -37.25 (1.08) | -55.79 (1.70) | -14.61 (0.03) | -62.60 (3.31) | 8.88 (0.88) ×10^-3 | -2.06 (2.48 × 10^-2) |

Table 3: **Log Bayes Factors**: Analysis of VAR series of US mini indexes as a function of data size. Average log Bayes Factors and standard deviation of log Bayes Factors over 20 independent Markov chains each of chain length 20,000.

| Rank \ T | 50     | 100    | 150    | 200    | 250    | 300    | 350    | 400    |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| r = 0   | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| r = 1   | 4.81 (1.00) | 3.14 (0.43) | 5.36 (1.06) | 5.92 (0.84) | 3.32 (0.75) | 1.30 (0.37) | 7.30 (0.59) | 3.10 (0.48) |
| r = 2   | -1.67 (12.78) | 3.66 (3.87) | -3.75 (3.04) | -1.83 (2.61) | -6.02 (3.22) | 0.14 (6.51) | -2.93 (1.96) | -7.73 (2.46) |
| r = 3   | -42.44 (12.38) | -48.58 (2.85) | -33.12 (0.14) | -100.42 (4.82) | -25.91(6.52 × 10^-2) | -10.33 (0.72) | -142.89 (3.31) | -195.47 (4.71) |

Table 4: **Log Bayes Factors**: Analysis of VAR series of US Bonds (5,10,30 Year Notes) as a function of data size. Average log Bayes Factors and standard deviation over 20 independent Markov chains each of chain length 20,000.
Figure 1: S&P 500, Dow Jones and Nasdaq mini Index daily close price data between 01-May-08 to 18-Sep-08. Left column plots represent scaled raw prices; Right plots represent difference data series.

Figure 2: 5, 10, 30 Year Notes - daily close price data between 01-May-08 to 18-Sep-08. Left column plots represent scaled raw prices; Right plots represent difference data series.
Figure 3: Empirical distribution of the Bayesian Model Averaging and Bayesian Model Order Selection, predictive performance for a combination of mini-index and bond data, taken over random intervals.