38

Bayesian Analysis of Cointegrated Vector Autoregressive Models

By

Katsuhiro Sugita

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Economics

Department of Economics

University of Warwick

March 2004

Contents

List of Figures

List of Tables

To My Wife, Maya

Acknowledgments

I am grateful to Mike Clements and Jeremy Smith for the continuous and patient guidance they have provided during the elaboration of this work.

This thesis has benefited from the comments and suggestions of participants in the workshop "Recent Advances in Bayesian Econometrics" in Marseilles, the Econometric Society's Far Eastern Meeting in Kobe, the ESRC Econometrics Study Group annual conference at Bristol, the European Economic Association annual congress in Lausanne, SCE conference at Yale, and Royal Economic Society's annual conference at Warwick, the Econometric Workshop at University of Warwick, seminars at Birkbeck, Essex, Hitotsubashi Universities. I wish particularly to thank Luc Bauwens, Richard Paap and Christpher Sims for their comments on earlier versions of some parts of this work. I also wish to thank John Chao and Richard Paap to provide me with codes that were used in writing this thesis.

Declaration

All the material in this thesis is my own work. The thesis has not been submitted for a degree at another university.

Summary

This thesis concerns econometric time series modelling of cointegrated multivariate systems using a Bayesian approach. The Bayesian approach has become increasingly attractive among researchers in the fields such as biology, though still only a relatively few econometricians use these techniques. Rather than theoretical aspects of Bayesian statistics or computational techniques, we illustrate how the Bayesian methods can be useful in analysing non-linear cointegration models.

In the last ten years, non-linear time series models, such as regime switching models, have become popular among applied econometricians to analyse the business cycles, policy evaluation in specific macroeconomic issues and forecasting. Cointegration analysis has been influenced by the non-linearity so that cointegration models that allow regime switching or structural breaks have been analysed by many econometricians. Unfortunately, these non-linear cointegration models tend to be complicated both in terms of estimation and testing.

We consider in this thesis a Bayesian approach to (i) a linear cointegration model, (ii) a cointegration model with Markov regime switching, and (iii) a cointegration model with multiple structural breaks, and show how easily we can analyse these models without any substantial modification.

Chapter 2 proposes a simple method for detecting cointegration rank using the Bayese factors, computed by the harmonic mean of the likelihood or Schwarz' Bayesian information criterion. Then we perform Monte Carlo simulations to compare three Bayesian methods (Phillips posterior information

criterion, Kleibergen and Paap method, and one proposed method) for the cointegration rank. Provided we have enough large sample size, the Phillips' posterior information criterion gives consistent results, while the results by Kleibergen and Paap method depends on the prior hyperparameters that we specify.

In Chapter 3, we develop the cointegration model that allows cointegration relationships to be switched on and off depending on the regime. Unlike the classical method that requires a two-step estimation, the Bayesian method provide a straightforward estimation and testing procedure.

In Chapter 4, we consider cointegration model with multiple structural breaks in the level, trend and error covariance. The more general model with breaks in both the adjustment term and the cointegrating vectors are also presented. To date, there is no research that deals with a cointegration model with unknown multiple structural breaks in any subset of the parameters.

Chapter 1

Introduction

1.1 Bayesian Analysis of Cointegration Models

Since the prominent papers by Granger (1981) and Engle and Granger (1987), testing for and estimating cointegrating regressions has become an integral part of the tools of the applied economic researchers. Many methods have been developed using either residual-based single equation or multivariate system to determine these cointegrating relationships. Most of these methods are done using the frequentist approach, based on the asymptotic properties. Among these classical methods, the Johansen's trace test and maximum eigenvalue test have been most widely used.

In contrast, the Bayesian approach to cointegration models have been developed by only a few econometricians, see, for example, Koop (1991 and 1994), DeJong (1992), Dorfman (1995), Kleibergen and van Dijk (1994), Geweke (1996), Bauwens and Lubrano (1996), Chao and Phillips (1999), Kleibergen and Paap (2003), Strachan (2003), Villani (2003), Strachan and

Inder (2004).

As Maddala and Kim (1998) have noted, the Bayesian approach to cointegration has advantages over the classical methods. Firstly, it gives us finite sample results, while most of the classical methods rely on the asymptotic distributions. Secondly it avoids pre-testing problem which arises in the classical methods. The pre-testing problem includes checking for unit roots for all variables in the model before undertaking the cointegration tests, and not knowing what effect these tests have on the significance levels used for the subsequent cointegration tests. Also, there is no definite answer to the question of what significance levels should be used for the unit root tests although it is conventional to use 5 percent and 1 percent significance levels. The pre-testing problem can be avoided using Bayesian methods because respective posterior probabilities of unit roots and stationarity are taken into account.

Despite these advantages of the Bayesian approach over the classical approach, it remains the case that the overwhelming majority of applied papers use the classical approach. There are several reasons for this observation. First, most econometricians rely on the classical method because they were not taught Bayesian econometrics. Until recently, there had not been available an appropriate textbook specialised in Bayesian econometric analysis. Zellner (1971) has been a standard econometric textbook, however, it does not cover the computational advances that have revolutionized Bayesian econometrics. Poirier (1995) focuses on the theoretical aspects on both Bayesian and frequentists approaches, but does not cover models other than a basic regression. Bauwens, Lubrano and Richard (1999) is the first

Bayesian textbook that focuses on time series econometrics, and Koop (2003) is the first general Bayesian econometrics textbook which treats broad topic within econometrics such as panel data regression models, limited dependent variable models, time series models and nonparametric and semiparametric methods. The second reason is that Bayesian approach involves heavy computation both analytically and numerically. But, with recent development of algorithms for the Markov Chain Monte Carlo (MCMC) methods and availability of faster computers has enabled us to undertake such complicated computation. The fourth reason is the lack of available computer packages for Bayesian techniques, meaning that one has to write ones own computer code if one wants to use Bayesian method. The final reason is that Bayesian method suffers from the controversy regarding to the choice of the prior density. It seems that there is still no consensus about choosing appropriate priors for the unit root regressions. For this topic, Kass and Wasserman (1996) provide a critical survey of the different methods of generating prior distributions.

This thesis is not concerned with theoretical aspects of Bayesian cointegration analysis nor discussion of choice of the prior, but concerned with how Bayesian methods can be applied to analyse various types of cointegration models such as cointegration models with Markov regime switching or with multiple structural breaks in level, trend and error covariance.

In classical methods, more complicated cointegration models such as nonlinear cointegration models require multiple steps for making inference on parameters of the models, which is not efficient compared with the Bayesian method. For example, in a Markov switching cointegration model where some parameters of the model are subject to Markov switching behavior, one has to estimate the parameters assuming the model is linear before unobserved regime variables are estimated, although the parameters are dependent on the regime variables. In the Bayesian analysis, both the parameters of the model and unknown regime variables are treated as random variables, and thus inference on the regime variables is based on a joint distribution. By using Gibbs sampling, both the parameters of the model and the unobserved regime variables are generated from appropriate conditional distributions.

As for the cointegration models with structural breaks, the Bayesian approach is much flexible and technically simpler than the classical methods as mentioned by Maddala and Kim (1998). For example, in a case of multiple structural breaks, Bai (1997) and Bai and Perron (1998) derived asymptotic distribution properties and proposed a method to detect the number of breaks in the coefficient of the regressors of the model. To use this method to analyse the model with breaks in trend and/or variance, one needs substantial modification while inference from the Bayesian approach is the same whether trend and/or variance is also subject to change. As in the case of Markov switching cointegration models, the Bayesian approach allows us to make inference on the break dates and other parameters of the cointegration models jointly using the Gibbs sampling. Another issue regarding to the models with structural breaks is that most classical methods focus on testing whether there is break or not (for example, Andrews (1993), Andrews and Ploberger (1994)). Bai and Perron (1998) discussed the problem of consistent estimation of the break points. The Bayesian approach gives consistent estimators and useful results such as uncertainty of the number and the locations of the break dates as the results of the Gibbs outputs of the posterior densities.

In this thesis we show how the Bayesian approach provides flexible and simple solutions for dealing with more complicated cointegration models with Markov regime switching in the level and the adjustment term and multiple structural breaks in the level, trend and error covariance.

1.2 Plan of the Thesis

In this section we provide a general outline of the thesis. The thesis consists of three chapters. Chapter 2 deals with linear cointegration models and evaluates three Bayesian testing methods for the cointegration rank using the Monte Carlo simulations. Chapter 3 is concerned with nonlinear cointegration models and applies the Bayesian method to evaluate a Markov switching cointegration model where the cointegrating relationships are subject to regime switching behavior using a discrete first order Markov process. Chapter 4 applies the Bayesian approach to analyse a cointegration model with multiple structural breaks in level, trend and error covariance.

A detailed outline of the three chapters are as follows. Chapter 2 overviews two existing Bayesian methods of analysing linear cointegration models, Phillips' posterior information criterion (PIC) and Kleibergen and Paap method (KP), and then considers a simple method of estimating the cointegration rank using the Bayes factors. Monte Carlo experiments are conducted to compare the three methods. Although the PIC enables us to select both the rank and the lag length jointly, we focus on the performance of the methods in selecting the cointegration rank, assuming the true lag length is known. We present two illustrative examples - Great ratios and PPP - to see how these three methods give different results.

Chapter 3 introduces a Bayesian approach to a Markov switching cointegration model that allows the cointegration relationships to be switched on and off depending on the regime. We also consider a less restrictive Markov switching cointegration model in which deviations from the long-run equilibrium are characterised by different rates of the adjustment depending upon the regimes. Unlike a classical method for nonlinear cointegration model that uses the cointegrating vector based on a linear cointegration model, the proposed Bayesian method allows for estimation of the cointegrating vector within a nonlinear framework conditional on the regime variables within the Gibbs sampling iteration. The Bayes factors are applied to test for Markov switching and model specifications. The PPP relationship between UK-US is investigated using the proposed model for illustration.

Chapter 4 investigates the issue of multiple structural breaks in the level, trend and error covariance of a cointegration model using a Bayesian approach. We take the number of cointegration relations as given and assume it is constant across the breaks, although it is easy to modify the procedure to allow the cointegration rank to change with the breaks. Estimation of the model is made possible by the use of the Gibbs sampler. The determination of the number of structural breaks is determined as sort of model selection using Bayes factors approximated by Schwarz's Bayesian information criterion from the data. The Bayesian method gives us the parameter uncertainty as results of the Gibbs sampling, so it gives uncertainty around break dates.

The model is applied to Japanese term structure data, and find that there is evidence of three structural breaks.

Chapter 5 summarises the main findings of this thesis and indicates directions for future research.

Chapter 2

Bayesian Cointegration Analysis

2.1 Introduction

In the past decade, the econometric literature on cointegration has grown markedly since Granger (1981) introduced the concept of cointegration and Engle and Granger (1987) presented the Error Correction Model representation and proposed a residual based test. Since then, many methods have been developed. For example, the FM-OLS procedure by Phillips and Hansen (1990), the dynamic OLS method by Saikkonen (1991), the nonlinear least squares by Phillips and Loretan (1991), the dynamic generalized least squares by Stock and Watson (1993). Among numerous procedures for estimation and testing for cointegration, Johansen's (1991) trace test and maximum eigenvalue test, based on canonical correlation in the system, have been most widely employed, and thus have been implemented in many econometrics packages such as *EViews, PcGive, Microfit,* and others.

Several researchers have proposed Bayesian inference in cointegrated VAR systems. Koop (1991) developed a Bayesian cointegration test using Monte Carlo integration techniques. He investigated the bivariate system of stock prices and dividends, tested for cointegration using posterior odds for hypotheses, and found that there is evidence to support that unit roots are not present in stock price and dividend and thus there is no cointegration relationship between the two series even if unit roots are assumed. DeJong (1992) developed a method for evaluating the co-integration inference over trend stationary alternatives, and examined cointegration relationship between consumption and income for the permanent income hypothesis. Dorfman (1995) used a posterior odds ratio test for cointegration on the number of nonstationary roots in the system, and tested for cointegration among the exchange rates. Koop (1994) proposed a method based on the number of nonstationary roots in a VAR system.

As for Bayesian cointegration analysis based on the framework of a vector error correction model, Kleibergen and van Dijk (1994) proposed using a Jeffrey's prior instead of a diffuse prior for the cointegrating vectors, since the marginal posteriors may be nonintegrable. Geweke (1996) developed general methods for Bayesian inference with noninformative reference priors in the reduced rank regression model. Bauwens and Lubrano (1996) reduced the VECM to a simple multivariate regression model to identify the parameters. The cointegrating rank is assumed to be known a priori, based on a theoretical economic model that defines equilibrium economic relations. If we are interested in identifying the cointegration rank, they suggest checking the plot of the posterior density of the eigenvalues of generated sample $\Pi'\Pi$, where Π denotes the long-run multiplier matrix, which are equal to the square of the singular values of Π. However, this informal visual inspection gives ambiguous results.[1] Bauwens, *et al* (1999) suggest using the trace test of Johansen, since "on the Bayesian side, the topic of selecting the cointegrating rank has not yet given very useful and convincing results"(p.283).

For more a formal Bayesian test for the cointegration rank, Kleibergen and Paap (2002) (KP, hereafter) proposed a method which uses a singular value decomposition of the unrestricted long-run multiplier matrix, Π, for identification of the cointegrating vectors and for Bayesian posterior odds analysis of the rank of Π. Chao and Phillips (1999) used the posterior information criterion (PIC, hereafter), proposed by Phillips and Ploberger (1994, 1996) and Phillips (1994a, 1994b, 1995, 1996), to select an appropriate model in terms of the rank and number of lags in the cointegrated VAR model.

Recent research by Strachan (2003) and Strachan and Inder (2004) criticised conventional prior with linear restrictions for the cointegrating vectors, and proposed a valid prior based on the cointegrating space. Strachan and van Dijk (2003b) applied this 'valid priors' to the VAR model. The choice of priors is also discussed by Strachan and van Dijk (2003a). Villani (2003) pointed out that the cointegration space is not an inner product space due to the well known non-identification of the cointegration vectors, and then proposed a Bayes estimator of the cointegration space that takes the curved geometry of the parameter space into account.

In this chapter we are interested in the performances of the two methods of KP and the PIC in the Monte Carlo simulations. We also introduce a

[1] Tsurumi and Wago (1996) use a highest-posterior-density-region (HPDR) test to Π , then derive the posterior pdfs for singular values to see whether 99% highest-posterior-density-interval (HPDI) contains zero.

simple method for determining the cointegration rank by Bayes factors. The method is very straightforward, that involves computing the Bayes factors for each possible rank, and then selecting rank which has the highest Bayes factor. The procedure for obtaining the posteriors has some similarities with Bauwens and Lubrano (1996) method with conventional priors with linear restrictions on the cointegrating vectors. Although Strachan (2003) criticised this prior for the cointegrating vectors as invalid, we follow the conventional prior in this thesis. We consider Strachan's method for the future research. While the method is not invariant with respect to the ordering of the variables in the VAR, it is able to determine the correct cointegrating rank. We conduct Monte Carlo simulations to compare this simple method with the KP method or PIC.

The plan of this chapter is as follows. We review the PIC and the KP method to detect the cointegration rank in Section 2.2. Section 2.3 presents a simple Bayesian method, specifies the prior densities, and derives the posterior densities for estimation of the cointegrated VAR systems. In Section 2.4 Bayes factors for determining the cointegration rank is introduced. Section 2.5 shows Monte Carlo simulations to compare the performance of the proposed method with the KP and the PIC method for determining the cointegration rank under different prior specifications. KP did not show any Monte Carlo simulations in their paper so that it is of interest to evaluate how the method performs compared with other methods. Although Chao and Phillips (1999) presented a small simulation study in their paper, the DGPs are limited to have zero or one rank with one to two lags in VAR. In Section 2.6, illustrative examples of the 'great ratios' (King *et al*, 1991) and purchasing power parity (PPP) are presented. Section 2.7 concludes.

2.2 Bayesian Approach to Cointegrated Multivariate Time Series Model - An Overview

We review briefly in this section cointegration tests by Phillips PIC and the KP method.

2.2.1 Posterior Information Criterion (PIC)

For model selection, the Akaike information criterion (AIC) or Schwarz's Bayesian information criterion (SBIC), which impose a penalty based on the dimension of the selected model, are most widely used (for example, to select the lag length in the model). Phillips and Ploberger (1994, 1996) and Phillips (1994a, 1994b, 1995) proposed an alternative criterion for model selection, called the posterior information criterion (PIC). The PIC explicitly depends on the data matrix, unlike both the Schwarz BIC and AIC which depends on the number of regressors and the residual variances. To select an appropriate model of the regression among different models $M = 1, \ldots, k$, $Y_M = X_M \beta_M + \epsilon$, we compute the PIC for all models $M = 1, \ldots, k$, and then select a model which has the lowest value of the PIC. The PIC is computed as follows:

$$\mathrm{PIC} = c_M \left| \frac{X_M' X_M}{\widehat{\sigma}_M^2} \right| \exp \left[-\frac{\widehat{\beta}_M' (X_M' X_M) \widehat{\beta}_M}{2\widehat{\sigma}_M} \right]^{1/2} \tag{2.1}$$

where c_M is a constant depending on M, the maximum number of regressors,

$\widehat{\sigma}_M^2$ is the maximum likelihood (ML) estimate of the error variance, and $\widehat{\beta}_M$ is the ML estimate of the coefficient vector.

Chao and Phillips (1999) extended the PIC to cointegrated models to select both the lag length $\widehat{p}$ and the number of rank $\widehat{r}$ jointly. They stressed this joint test since the performance of cointegration test such as Johansen (1992) can be adversely affected by lag misspecification as shown by Toda and Phillips (1994).

Consider the n-dimensional vector autoregressive process of order $p+1$

$$Y_t = \Phi(L)Y_{t-1} + \epsilon_t \tag{2.2}$$

where $\Phi(L) = \Sigma_{i=1}^{p+1}\Phi_i L^{i-1}$. Eq. (2.2) can be written in vector error correction model (VECM) representation as

$$\Delta Y_t = \Phi^*(L)\Delta Y_{t-1} + \Pi_* Y_{t-1} + \epsilon_t \tag{2.3}$$

where $\Pi_* = \Phi(1) - I_n = \alpha\beta'$ with α and β are $n \times r$ matrices, and $\Phi^*(L) = \Sigma_{i=1}^{p+1}\Phi_i^* L^{i-1}$ with $\Phi_i^* = -\Sigma_{m=i+1}^{p+1}\Phi_m$, $i = 1, \ldots, p$.

Let $Y = [Y_1, \ldots, Y_T]'$, $Y_{-1} = [Y_0, \ldots, Y_{t-1}]'$, $\Delta Y = [\Delta Y_1, \ldots, \Delta Y_T]'$ and $W(p) = [W_1(p), \ldots, W_T(p)]'$ with $W_t(p) = \left[\Delta Y_{t-1}', \ldots, \Delta Y_{t-p}'\right]'$, $W(\widetilde{p}) = \begin{bmatrix} W(p) & W(p^*) \end{bmatrix}$ where $W(p)$ contains the first np columns and $W(p^*)$ contains the last $n(\widetilde{p} - p)$ columns of the $T \times n\widetilde{p}$ matrix $W(\widetilde{p})$. $F(r)$ is the $n \times (n-r)$ matrix defined as $F(r) = \begin{bmatrix} 0 & I_{n-r} \end{bmatrix}$. In addition, let $X = [\Delta Y, Y_{-1}, W(p), W(p^*)]$ and $S = X'X$ and write S in partitioned form as:

$$\begin{aligned} S &= \begin{bmatrix} \Delta Y'\Delta Y & \Delta Y'Y_{-1} & \Delta Y'W(p) & \Delta Y'W(p^*) \\ Y_{-1}'\Delta Y & Y_{-1}'Y_{-1} & Y_{-1}'W(p) & Y_{-1}'W(p^*) \\ W(p)'\Delta Y & W(p)'Y_{-1} & W(p)'W(p) & W(p)'W(p^*) \\ W(p^*)'\Delta Y & W(p^*)'Y_{-1} & W(p^*)'W(p) & W(p^*)'W(p^*) \end{bmatrix} \\ &= \begin{bmatrix} S_{\Delta\Delta} & S_{\Delta y} & S_{\Delta p} & S_{\Delta p^*} \\ S_{y\Delta} & S_{yy} & S_{yp} & S_{yp^*} \\ S_{p\Delta} & S_{py} & S_{pp} & S_{pp^*} \\ S_{p^*\Delta} & S_{p^*y} & S_{p^*p} & S_{p^*p^*} \end{bmatrix} \\ &= \begin{bmatrix} S_{\Delta\Delta} & S_{\Delta y} & S_{\Delta\widetilde{p}} \\ S_{y\Delta} & S_{yy} & S_{y\widetilde{p}} \\ S_{\widetilde{p}\Delta} & S_{\widetilde{p}y} & S_{\widetilde{p}\widetilde{p}} \end{bmatrix} \end{aligned}$$

Define $S_{ij.k} = S_{ij} - S_{ik}S_{kk}^{-1}S_{kj}$ for $i,j = \Delta, y$ and $k = p, \widetilde{p}$, and $S_{ij.k.l} = S_{ij.k} - S_{il.k}S_{ll.k}^{-1}S_{lj.k}$ for $i.j = \Delta, p^*$ and $k, l = y, p, \widetilde{p}$.

To estimate the cointegrating rank, r, and the lag length, p, jointly, we select $(\widehat{p}, \widehat{r})$ as follows:

$$(\widehat{p}, \widehat{r}) = \arg\min \mathrm{PIC}(p, r)$$

where

$$\mathrm{PIC}(p,r) = \exp\left\{\frac{1}{2}\mathrm{tr}\left[\widehat{\Sigma}^{-1}\left(\widetilde{\Pi}_*(p,r) - \widehat{\Pi}_*(p)\right) S_{yy.p}\left(\widetilde{\Pi}_*(p,r) - \widehat{\Pi}_*(p)\right)'\right]\right\}$$

$$\times \exp\left\{\frac{1}{2}\mathrm{tr}\left[\widehat{\Sigma}^{-1}\widehat{\Phi}^*(p^*)S_{p^*p^*.y.p}\widehat{\Phi}^*(p^*)'\right]\right\}\left[\left|\widehat{\Sigma}^{-1}\otimes S_{pp}\right|^{1/2} / \left|\widehat{\Sigma}^{-1}\otimes S_{\overline{p}\overline{p}}\right|^{1/2}\right]$$
$$\times\left[\left|\widetilde{H}(p,r)\left(\widehat{\Sigma}^{-1}\otimes S_{yy.p}\right)\widetilde{H}(p,r)'\right|^{1/2} / \left|\widehat{\Sigma}^{-1}\otimes S_{yy.\overline{p}}\right|^{1/2}\right] \tag{2.4}$$

where $\widetilde{\Pi}_*(p,r) = \left[\widehat{\alpha}(p,r), \widehat{\alpha}(p,r)\widehat{\overline{\beta}}(p,r)'\right]$ with $\widehat{\alpha}(p,r)$ and $\widehat{\overline{\beta}}(p,r)$ are the maximum likelihood estimators of the parameters α and $\overline{\beta}$ when the cointegrating rank is assumed to be r and the number of lags is assumed to be p, $\widehat{\Pi}_*(p) = S_{\Delta y.p}S_{yy.p}^{-1}$ and $\widehat{\Pi}^*(p^*) = S_{\Delta p^*.y.p}S^*_{p^*p^*.y.p}$, $\widehat{\Sigma}$ is the maximum likelihood estimator of Σ, and the $(2nr - r^2)\times n^2$ matrix $\widetilde{H}(p,r) = \left[(\widehat{\alpha}(p,r)'\otimes F(r)')', (I_n \otimes (I_r\widehat{\overline{\beta}}(p,r)'))'\right]'$.

Chao and Phillips (1999) criticized Johansen's sequential procedure of testing the cointegrated rank from the subhypothesis $r = 0$ onwards as the procedure does not yield a consistent estimator of the cointegrating rank. Another advantage to the PIC is that the penalty function of the PIC takes into account not only the number of parameters (like AIC and SBC) but also the nonstationarity of the regressors associated with some of the parameters.

However, the procedure is not completely Bayesian because some parameters rely on the maximum likelihood estimators. Also, unlike Bayesian posterior odds analysis, the PIC does not provide uncertainty among models that we consider (see Phillips, 1995, the comments, and Phillips' reply).

2.2.2 Kleibergen and Paap (2002)

Kleibergen and Paap (2002) proposed a Bayesian method for analysing rank reduction of the long-run multiplier matrix in a vector autoregressive model by using a singular value decomposition to construct a parameter that reflects the presence of rank reduction.

Suppose we are considering the VECM of the form:

$$\Delta X_t = \mu + \Pi' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t, \tag{2.5}$$

KP decomposed the long-run multiplier, Π, as follows:

$$\Pi = \beta\alpha + \beta_\perp \lambda \alpha_\perp = \begin{bmatrix} \beta & \beta_\perp \end{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha_\perp \end{bmatrix} \tag{2.6}$$

where $\alpha_\perp$ and $\beta_\perp$ are specified such that $\alpha_\perp \alpha' \equiv 0$ with $\alpha_\perp \alpha'_\perp \equiv I_{n-r}$ and $\beta'_\perp \beta \equiv 0$ with $\beta'_\perp \beta_\perp \equiv I_{n-r}$. When $\lambda = 0$, the long-run multiplier Π shows rank reduction and the model has some cointegrating vectors. By applying the singular value decomposition on Π, we have $\Pi = USV'$, where U and V are $n \times n$ orthonormal matrices with $U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$ and $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$, and S is an $n \times n$ diagonal matrix containing the non-negative singular values with decreasing order such that $S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$, U_{11}, S_1, and V_{11} are $r \times r$ matrices, U_{22}, S_2 and V_{22} are $(n-r) \times (n-r)$ matrices, U_{21} and V_{21} are $(n-r) \times r$ matrices, and U_{12} and V_{12} $r \times (n-r)$ matrices. Then, we have $\alpha = U_{11} S_{11} [V_{11}, V_{21}]'$, $\beta = -U_{21} U_{11}^{-1}$, $\lambda = (U'_{22} U_{22})^{-1/2} U_{22} S_2 V'_{22} (V_{22} V'_{22})^{-1/2}$. The importance sampler with Chen's importance weights (1994) or the Metropolis-Hastings (M-H) algorithm (Metropolis *et al* (1953) and Hastings (1970)) can be implemented to generate posterior output instead of using the standard Gibbs sampling because the full conditional posterior distributions are of unknown type. Testing the cointegration rank is done by the posterior odds, and the Bayes factors are computed using the Savage-Dickey density ratio (Dickey, 1971) by imposing $\lambda = 0$ as the null hypothesis.

KP chose the conjugate priors, the inverted Wishart for the covariance matrix Σ and the matric variate normal for Π conditional on Σ, and the g-prior of Zellner (1986) for the prior covariance for Π. The joint prior on the parameters in the cointegration model is obtained by putting the matrix reflecting the presence of rank reduction ($\lambda = 0$) such that $p(\Sigma, \alpha, \beta) \propto p(\Sigma, \Pi) \mid_{\Pi=\beta\alpha} |J(\Pi, (\alpha, \lambda, \beta)|_{\lambda=0}$where $|J(.)|_{\lambda=0}$ denotes the Jacobian transformation evaluated in $\lambda = 0$. In case of diffuse (non-informative) prior specification, KP asserted that by taking an appropriate prior height $(2\pi)^{-1/2(n-r)^2}$, the Bayes factor is equivalent to the PIC.

2.3 Bayesian Inference in Cointegration Analysis

2.3.1 Statistical Model

In this section we present a simple Bayesian analysis of cointegration, extending Bauwens and Lubrano (1996). Let X_t denote an $I(1)$ vector of n-dimensional time series with r linear cointegrating relations, then the unrestricted VECM representation with deterministic trend is:

$$\Delta X_t = \mu + \gamma t + \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{2.7}$$

where $t = p, p+1, \ldots, T$, p is the number of lags in VAR, and the errors, ε_t, are assumed $N(0, \sum)$ and independent over time. μ, γ, ε, Ψ, Σ, α, and β are parameters of dimensions $n \times 1$, $n \times 1$, $n \times n$, $n \times n$, $n \times r$, and $n \times r$, respectively.

Equation (2.7) can be rewritten in matrix format as:

$$Y = X\Gamma + Z\beta'\alpha' + E = WB + E \tag{2.8}$$

where

$$Y = \begin{bmatrix} \Delta X_p' \\ \Delta X_{p+1}' \\ \vdots \\ \Delta X_T' \end{bmatrix}, Z = \begin{bmatrix} X_{p-1}' \\ X_p' \\ \vdots \\ X_{T-1}' \end{bmatrix}, E = \begin{bmatrix} \varepsilon_p' \\ \varepsilon_{p+1}' \\ \vdots \\ \varepsilon_T' \end{bmatrix}, \Gamma = \begin{bmatrix} \mu' \\ \gamma' \\ \Psi_1' \\ \vdots \\ \Psi_{p-1}' \end{bmatrix},$$

$$X = \begin{bmatrix} 1 & p & \Delta X_{p-1}' & \cdots & \Delta X_1' \\ 1 & p+1 & \Delta X_p' & \cdots & \Delta X_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & T & \Delta X_{T-1}' & \cdots & \Delta X_{T-p+1}' \end{bmatrix}, W = \begin{bmatrix} X & Z\beta' \end{bmatrix},$$

$$B = \begin{bmatrix} \Gamma \\ \alpha' \end{bmatrix}.$$

Let m be the number of rows of Y, so that $m = T - p + 1$, then X is $m \times (2+n(p-1))$, Γ $((2+n(p-1)) \times n)$, W $(m \times k)$, where $k = 2+n(p-1)+r$, and B $(k \times n)$. Thus, equation (2.8) represents the multivariate regression format of (2.7). This representation is a starting point. We then describe the prior and likelihood specifications in order to derive posteriors.

2.3.2 Prior and Posterior Distributions

In this subsection, we consider a Bayesian estimation of the vector error correction models presented in (2.7). The conjugate prior density for B conditional on covariance Σ follows a matrix-variate normal distribution with covariance matrix $\Sigma \otimes A^{-1}$ of the form

$$p(B \mid \Sigma) \propto |\Sigma|^{-k/2}|A|^{n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(B-P)' A (B-P)\right\}\right] \tag{2.9}$$

where A is $(k \times k)$ PDS and P $(k \times n)$, $k = n(p-1)+r+1$ (the number of columns in W).

For the prior density of the covariance Σ in (2.8), we can assign an inverted Wishart

$$p(\Sigma) \propto |S|^{h/2}|\Sigma|^{-(h+n+1)} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}S\right)\right\} \tag{2.10}$$

where h represents the degrees of freedom, S an $n \times n$ PDS. Instead of above priors, if we do not want to impose an informative prior for Σ, we can opt diffuse prior for Σ as $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$.

The prior for β can be given as a matrix-variate normal

$$\pi(\beta) \propto |Q|^{-n/2} |H|^{r/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{Q^{-1}\left(\beta-\overline{\beta}\right)' H \left(\beta-\overline{\beta}\right)\right\}\right] \tag{2.11}$$

where $\overline{\beta}$ is a prior mean of β, Q is $r \times r$ PDS, H is $n \times n$ PDS. Note that r^2 restrictions for identification are imposed on β, for example, $\beta' = \begin{pmatrix} I_r & \beta_\star' \end{pmatrix}$,[2] where $\beta_\star$ is $(n-r) \times r$ unrestricted matrix. If we assign r^2 restrictions on β as I_r, then only a part of β, $\beta_\star$, follows a matrix-variate normal.

If we assume that B and Σ are independent of β, then the joint prior of the parameters in (2.8) is $p(B,\beta,\Sigma) \propto p(B|\Sigma)p(\beta)p(\Sigma)$ and thus can be derived as

$$p(B,\Sigma,\beta) \propto \pi(\beta)\,|A|^{n/2}|\Sigma|^{-\frac{k+h+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S+(B-P)'A(B-P)\right]\right\}\right] \tag{2.12}$$

To derive the conditional posterior distributions, we need to derive the likelihood functions. The likelihood function for B, Σ, and β is given by:

$$\begin{aligned} L\left(Y \mid B,\Sigma,\beta\right) &\propto |\Sigma|^{-t/2} \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left(Y-WB\right)'\left(Y-WB\right)\right\}\right] \\ &\propto |\Sigma|^{-t/2} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left\{\widehat{S}+(B-\widehat{B})'W'W(B-\widehat{B})\right\}\right]\right\} \end{aligned} \tag{2.13}$$

where $\widehat{B} = (W'W)^{-1}W'Y$, and $\widehat{S} = (Y-W\widehat{B})'(Y-W\widehat{B})$.

Next we derive the posteriors from the priors and the likelihood function specified above. The joint posterior distribution for the conjugate priors for

[2]The restrictions imposed on β need not to be I_r but can be any r^2 restrictions. See Bauwens and Lubrano (1996, page 14)

B, Σ and β is proportional to the joint prior (2.12) times the likelihood function (2.13), thus we have

$$
\begin{aligned}
& p(B,\Sigma,\beta \mid Y) \propto p(B,\Sigma,\beta)\, L(Y \mid B,\Sigma,\beta) \\
\propto\; & \pi(\beta)\, |A|^{\frac{n}{2}} |\Sigma|^{-(t+h+k+n+1)/2} \\
& \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S+(B-P)'A(B-P)+\widehat{S}+(B-\widehat{B})'W'W(B-\widehat{B}\right]\right\}\right] \\
\propto\; & \pi(\beta)\, |\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S+\widehat{S}+(P-\widehat{B})'[A^{-1}+(W'W)^{-1}]^{-1}(P-\widehat{B})\right.\right.\right. \\
& \left.\left.\left. +(B-B_\star)'A_\star(B-B_\star)\right]\right\}\right] \\
=\; & \pi(\beta)\, |\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S_\star+(B-B_\star)'A_\star(B-B_\star)\right]\right\}\right] \qquad (2.14)
\end{aligned}
$$

where $c = t+k+h+n+1$, $A_\star = A+W'W$, $B_\star = (A+W'W)^{-1}(AP+W'W\widehat{B})$, and $S_\star = S+\widehat{S}+(P-\widehat{B})'[A^{-1}+(W'W)^{-1}]^{-1}(P-\widehat{B})$.

From (2.14), the conditional posterior of Σ is derived as an inverted Wishart distribution, and the conditional posterior of B as a matrix-variate normal density with covariance, $\Sigma \otimes A_\star^{-1}$, that is,

$$
p(\Sigma \mid \beta, Y) \propto |S_\star|^{t/2} |\Sigma|^{-(t+h+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}S_\star\right)\right] \qquad (2.15)
$$

$$
p(B \mid \Sigma,\beta,Y) \propto |A_\star|^{n/2} |\Sigma|^{-k/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(B-B_\star)'A_\star(B-B_\star)\right\}\right] \qquad (2.16)
$$

Thus, by multiplying (2.15) and (2.16), and integrating with respect to Σ, we obtain the posterior density of B conditional on β, which is a matrix-variate Student-t form,

$$p(B \mid \beta, Y) \propto |S_\star|^{t/2}|A_\star|^{n/2}|S_\star + (B - B_\star)'A_\star(B - B_\star)|^{-(t+k)/2} \tag{2.17}$$

The joint posterior of B and β can be derived by integrating (2.14) with respect to Σ,

$$p(B, \beta \mid Y) \propto \pi(\beta) \, |S_\star + (B - B_\star)'A_\star(B - B_\star)|^{-(t+h+k+1)/2} \tag{2.18}$$

By integrating (2.18) with respect to B we obtain the posterior density of the cointegrating vector β,

$$p(\beta \mid Y) \propto \pi(\beta) \mid S_\star \mid^{-(t+h+1)/2} \mid A_\star \mid^{-n/2} \tag{2.19}$$

The properties of (2.19) are not known, so that we have to resort to numerical integration techniques, as Bauwens and Lubrano (1996) used importance sampling to compute poly-t posterior results of the parameters. Other feasible methods are the Metropolis-Hastings algorithm and the Griddy-Gibbs sampling. The Metropolis-Hastings[3] algorithm requires the assignment of a good approximating function, the *candidate-generating function*, to the posterior to draw random numbers, as importance sampling requires the impor-

[3]For more details, consult Chen, *et al* (2000), Evans and Swartz (2000). For a tutorial for the M-H algorithm, see Chib and Greenberg (1995).

tance function. Since the Griddy-Gibbs sampling method does not require such an approximation, we employ the Griddy-Gibbs sampler for estimation of the cointegrating vector (Bauwens and Giot (1998) used this sampler for the estimation of two cointegrating vectors).

2.3.3 The Griddy-Gibbs Sampler

The Griddy-Gibbs sampler, proposed by Ritter and Tanner (1992), approximates the true cdf of each conditional distribution by a piecewise linear function and then samples from the approximations. This sampler can be implemented when the conditional posterior density is unknown to the researcher. The disadvantage of this sampling method is that the results are depending upon how we assign the range and the number of the grid for the parameter. The range should be chosen so that the generated numbers are not truncated. Another disadvantage is that this sampler demands more computing time than other algorithms. The advantage of using this sampler over the importance sampler or the Metropolis-Hastings algorithm is that researcher does not have to provide an approximation of the function. The procedure for implementing the Griddy-Gibbs sampler is as following:

1. Before we begin the chain, we must choose the range of the grid and the number of the grid. The range should be chosen so that the generated numbers are not truncated.

2. Let $\text{vec}(\beta)' = (\beta_1, \beta_2, \ldots, \beta_m)$. With an arbitrary starting value (within the upper and the lower bound of the grid), compute $f(\beta_1|\beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$, where i denotes the i-th loop, over the grid $(\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,U})$, where

$\beta_{1,1}$ is the lower bound of the grid of β_1, and $\beta_{1,U}$ is the upper bound of the grid of β_1.

3. Compute the values $G = (0, \Phi_2, \Phi_3, \ldots, \Phi_U)$ where

$$\Phi_j = \int_{\beta_{1,1}}^{\beta_{1,j}} f(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y) d\beta_1$$
$$j = 2, \ldots, U$$

4. Compute the normalized pdf values $G_\zeta = G_j / \Phi_U$ of $\zeta(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$.

5. Draw the random numbers from the uniform density with the lower bound as zeros and the upper bound as Φ_U and invert cdf G by numerical interpolation to obtain a draw β_1^i from $\zeta(\beta_1 | \beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$.

6. Repeat steps 2-5 for $\beta_2, \ldots, \beta_m$.

7. Set $i = i + 1$ (increment i by 1) and go to step 2.

Note that integration at the step 3 can be done by the deterministic approximation such as the Simpson's rule or the Trapezoidal rule.

2.4 Bayes Factors for Cointegration Tests

This section introduces the computation of the Bayes factors to determine the cointegration rank. The Bayes factor, which is defined as the ratio of the marginal likelihood of the null and the alternative hypotheses, has been used for model selection. Bayes factors can be used to construct posterior probabilities for all models that seem plausible. In classical hypothesis testing, one model represents the truth and the test is based on a pairwise comparison with the alternative. For a detailed discussion of the advantages of Bayesian methods, see Koop and Potter (1999). Kass and Raftery (1995) provide an excellent survey of the Bayes factor.

Suppose, with data Y and the likelihood functions with the parameters Θ, there are two hypotheses H_0 and H_1. The Bayes factor BF_{01} is defined as follows:

$$\begin{aligned} BF_{01} &= \frac{\Pr(Y|H_0)}{\Pr(Y|H_1)} \\ &= \frac{\int p(\Theta|H_0)L(Y|\Theta, H_0)d\Theta}{\int p(\Theta|H_1)L(Y|\Theta, H_1)d\Theta} \end{aligned} \tag{2.20}$$

With the prior odds, defined as $\Pr(H_0)/\Pr(H_1)$, we can compute the posterior odds, which are

$$\text{PosteriorOdds}_{01} = \frac{\Pr(H_0|Y)}{\Pr(H_1|Y)} = \frac{\Pr(Y|H_0)}{\Pr(Y|H_1)} \cdot \frac{\Pr(H_0)}{\Pr(H_1)} \tag{2.21}$$

When several models are being considered, the posterior odds yield the posterior probabilities. Suppose q models with $H_0, H_1, \ldots, H_{q-1}$ are being considered, and each of the hypotheses $H_1, H_2, \ldots, H_{q-1}$ is compared with H_0. Then the posterior probability for model i under H_i is

$$\Pr(H_i|Y) = \frac{\text{PosteriorOdds}_{i0}}{\sum_{j=0}^{q-1} \text{PosteriorOdds}_{j0}} \tag{2.22}$$

where $\text{PosteriorOdds}_{00}$ is defined to be 1. These posterior probabilities are used to select the cointegrating rank, model selection, or as weights for fore-

Table 2.1: Evaluating Bayes Factors

This table was reproduced from Kass and Raftery (1995)

BF	Evidence against H_0
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
> 150	Very strong

casting. A rule of thumb for interpreting the magnitude of a Bayes factor provided by Kass and Raftery (1995) is reproduced in Table 2.1 for convenience.

There are several methods to compute the Bayes factors given in (2.20). For example, the Laplace approximation method (Tierney and Kadane, 1986), or using numerical integration techniques such as importance sampling (Geweke, 1989) or the Metropolis-Hastings algorithm. See Kass and Raftery (1995) for details. Chib (1995) proposes a simple approach to compute the marginal likelihood from the Gibbs output. Newton and Raftery (1994) suggested using the posterior density $p(\theta \mid Y)$ as the importance function because samples from the posterior density arise directly from the Gibbs sampler, so that the marginal likelihood for model j (M_j) can be simplified to the harmonic mean of the likelihood as:

$$\Pr(Y|M_j) = \left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{L\left(\theta_j^{(k)} \mid Y; M_j\right)} \right]^{-1} \tag{2.23}$$

where $\theta^{(k)}$, $k = 1, \ldots, N$, are sample draws from the Gibbs sampler. An alternative approach to derive the Bayes factor is using the Schwarz BIC as

Yao (1988) and Liu *et al* (1997) suggested:

$$\mathrm{BIC}_j = -2\ln \mathfrak{L}\left(\widehat{\theta}_j \mid Y; M_j\right) + q_j \ln(t) \tag{2.24}$$

where $\mathfrak{L}\left(\widehat{\theta}_j \mid Y; \mathrm{M}_j\right)$ denotes the likelihood function under the model j; q_j denotes the total number of estimated parameters in the model j; M_j denotes the model indicator for model j. The likelihood function $\mathfrak{L}\left(\widehat{\theta}_j \mid Y; \mathrm{M}_j\right)$ is evaluated at $\widehat{\theta}_j$, the posterior means of the parameters for model j. The Bayes factor for model k against model j can be approximated by

$$BF_{jk} \simeq \exp\left[0.5\left(\mathrm{BIC}_j - \mathrm{BIC}_k\right)\right] \tag{2.25}$$

In this chapter we use two algorithms in (2.23) and (2.25) to detect the rank in cointegrated VAR model. Note that our method is not invariant with respect to ordering of the variables in the VAR, and thus the values of Bayes factors depend on the ordering, although, the values should reflect the correct rank.

2.5 Monte Carlo Simulation

To illustrate the performance of Bayesian tests for the rank of cointegration described in the previous section (2.3 - 2.4), we perform some Monte Carlo simulations. The data generating processes (DGPs) consist of a four-variable VAR with an intercept term having various number of cointegrating vectors (0, 1, 2, 3 and 4) as following:

1. $(r = 0)$ $\triangle y_t = \mu + e_t$

2. $(r=1)$ $\triangle y_t = \mu + \begin{bmatrix} -0.2 \\ -0.2 \\ -0.2 \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \end{bmatrix} y_{t-1} + e_t$

3. $(r=2)$ $\triangle y_t = \mu + \begin{bmatrix} -0.2 & -0.2 \\ 0.2 & -0.2 \\ 0.2 & 0.2 \\ -0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} y_{t-1} + e_t$

4. $(r=3)$ $\triangle y_t = \mu + \begin{bmatrix} -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} y_{t-1} + e_t$

5. $(r=4)$ $\triangle y_t = \mu + \begin{bmatrix} -0.2 & -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & 0.2 & -0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} y_{t-1} + e_t$

where $\mu = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}'$, and $e_t \sim NID(0, I_4)$.

We demonstrate the performance of the Bayes factors for determining the rank considering DGPs with different true rank. We undertake 1000 replications. The sample size t is 50, 100 and 200. We consider a VAR(1) model with a constant term throughout the experiments. As noted in the previous section, since the testing procedure presented in the previous section depends upon the chosen ordering of the variables in the VAR, the ordering of the individual series in y is changed randomly during the simulation experiment.

In these experiments, we also run the Monte Carlo simulations for the PIC and the KP method described in Section 2.2. The PIC does not depend on the prior distributions. Note that the PIC does not provide any posterior probability so that interpretation from the Monte Carlo simulations is not the same as from the other Bayesian methods.

The prior parameter specifications for the natural conjugate priors are as follows: $P = 0$ and $A = I_k/1000$ in (2.9), $\overline{\beta} = \beta$, $Q = I_n$, $H = I_3/1000$ in (2.11), $S = I_4/1000$ in (2.10) to ensure fairly large variance for representing prior ignorance. For the KP method, we assign $\theta = 1$ and 0.01 in $\theta(X'X)/t$, which is the prior variance of Π and g-prior of Zellner (1986), to see how this prior specification affects the results because this method uses the Savage-Dickey density ratio to compute the Bayes factor and thus it is very sensitive in choosing the prior parameters, while the Bayes factors by the methods in (2.25) and (2.23) are insensitive in the hyperparameters. Note that a smaller value of θ implies less prior information.

Table 2.2 - 2.4 summarizes the results of Monte Carlo simulation for sample sizes t is 50, 100 and 200 respectively. The values in the columns are the average posterior probabilities from 1,000 iterations for each true rank. For each iteration, the Griddy-Gibbs sampling is performed with 5,000 draws and the first 1,000 discarded for computing the posterior odds by the method we propose in this paper (labeled BF^1 and BF^2). The BF^1 is computed using the Schwarz BIC given in (2.25), and the BF^2 is obtained by using the harmonic mean of the likelihood given in (2.23). The column labeled as $\text{KP}^1$is the average posterior probabilities when $\theta = 1$, and KP^2 is when $\theta = 0.01$. The column labeled as KP^3 is when the diffuse priors are

selected. The column labeled as PIC is not the average posterior probability because the PIC does not offer posterior odds (model uncertainty) so that each elements is the frequency that each rank is chosen.

Tables 2.2 - 2.4 show that the PIC tends to select lower rank than the true rank especially when the sample size is 50 and 100. For example, with $t = 50$ and full rank, the PIC selects a correct rank with only 13.6 per cent. However, with $t = 200$, the PIC shows the best performance in our simulations among all methods we consider. This PIC's sample size sensitivity would be caused by the fact that the criterion uses the maximum likelihood estimators.

For the KP method, it is clear that the method is quite sensitive in the choice of the prior hyperparameter θ. The lower θ, the method tends to choose lower rank. The method with diffuse priors shows much better performance in our simulations, although it performs as poorly as the PIC when the sample size is small.

BF^1 also tends to select lower rank than the true rank when the sample size is small. This method also performs worse than BF^2 when the sample size is small. However, BF^2 shows slightly better performance than BF^1.

In summary, we see that the method with (2.25) would be good choice when the sample size is small. The KP method with informative priors is not useful unless we have strong prior knowledge. When the sample size is large, the PIC, the KP method with smaller value of hyperparameter θ or the KP method with diffuse prior can be good choices.

Table 2.2: Monte Carlo Results: $t = 50$: Average Posterior Probabilities

True rank	rank r	PIC	KP^1	KP^2	KP^3	BF^1	BF^2
$r = 0$	0	0.924	0.315	1.000	0.914	1.000	0.887
	1	0.078	0.288	0.000	0.085	0.000	0.109
	2	0.001	0.197	0.000	0.001	0.000	0.004
	3	0.000	0.122	0.000	0.000	0.000	0.000
	4	0.000	0.079	0.000	0.000	0.000	0.000
$r = 1$	0	0.267	0.009	0.982	0.296	0.806	0.222
	1	0.702	0.299	0.018	0.674	0.193	0.715
	2	0.030	0.307	0.000	0.026	0.000	0.062
	3	0.001	0.229	0.000	0.003	0.000	0.000
	4	0.000	0.156	0.000	0.000	0.000	0.000
$r = 2$	0	0.001	0.000	0.244	0.002	0.015	0.000
	1	0.329	0.033	0.573	0.644	0.168	0.135
	2	0.630	0.181	0.167	0.334	0.811	0.811
	3	0.038	0.450	0.013	0.017	0.003	0.026
	4	0.001	0.336	0.003	0.003	0.004	0.029
$r = 3$	0	0.003	0.000	0.325	0.001	0.007	0.000
	1	0.213	0.002	0.433	0.382	0.300	0.169
	2	0.501	0.121	0.211	0.457	0.386	0.287
	3	0.276	0.447	0.023	0.117	0.295	0.532
	4	0.007	0.430	0.008	0.043	0.015	0.011
$r = 4$	0	0.001	0.000	0.227	0.000	0.003	0.000
	1	0.188	0.001	0.473	0.260	0.307	0.023
	2	0.439	0.056	0.226	0.436	0.283	0.092
	3	0.236	0.262	0.038	0.165	0.105	0.182
	4	0.136	0.681	0.036	0.139	0.302	0.702

Note:
PIC: Posterior Information Criterion
KP^1: Kleibergen and Paap method with $\theta = 1.00$
KP^2: Kleibergen and Paap method with $\theta = 0.01$
KP^3: Kleibergen and Paap method with diffuse prior
BF^1: uses the Schwarz BIC given in (2.25)
BF^2: uses the harmonic mean of the likelihood given in (2.23)

Johansen: the numbers are p-values by Johansen's trace test

Table 2.3: Monte Carlo Results: $t = 100$: Average Posterior Probabilities

True rank	rank r	PIC	KP^1	KP^2	KP^3	BF^1	BF^2
$r = 0$	0	0.995	0.688	1.000	0.997	1.000	0.924
	1	0.005	0.213	0.000	0.003	0.000	0.076
	2	0.000	0.065	0.000	0.000	0.000	0.000
	3	0.000	0.023	0.000	0.000	0.000	0.000
	4	0.000	0.011	0.000	0.000	0.000	0.000
$r = 1$	0	0.011	0.000	0.468	0.038	0.303	0.058
	1	0.984	0.610	0.532	0.960	0.689	0.904
	2	0.005	0.263	0.000	0.003	0.009	0.038
	3	0.000	0.091	0.000	0.000	0.000	0.000
	4	0.000	0.036	0.000	0.000	0.000	0.000
$r = 2$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.020	0.006	0.165	0.198	0.002	0.020
	2	0.974	0.319	0.782	0.780	0.981	0.904
	3	0.006	0.414	0.043	0.021	0.016	0.074
	4	0.000	0.262	0.011	0.001	0.000	0.001
$r = 3$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.006	0.000	0.134	0.032	0.083	0.058
	2	0.487	0.032	0.624	0.648	0.374	0.222
	3	0.505	0.601	0.229	0.286	0.543	0.720
	4	0.002	0.365	0.013	0.034	0.000	0.000
$r = 4$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.003	0.000	0.100	0.015	0.021	0.000
	2	0.278	0.005	0.445	0.364	0.276	0.037
	3	0.367	0.093	0.138	0.199	0.158	0.177
	4	0.353	0.902	0.318	0.422	0.545	0.787

Note:
PIC: Posterior Information Criterion
KP^1: Kleibergen and Paap method with $\theta = 1.00$
KP^2: Kleibergen and Paap method with $\theta = 0.01$
KP^3: Kleibergen and Paap method with diffuse prior
BF^1: uses the Schwarz BIC given in (2.25)
BF^2: uses the harmonic mean of the likelihood given in (2.23)

Johansen: the numbers are p-values by Johansen's trace test

Table 2.4: Monte Carlo Results: $t = 200$: Average Posterior Probabilities

True rank	rank r	PIC	KP^1	KP^2	KP^3	BF^1	BF^2
$r = 0$	0	0.999	0.897	1.000	1.000	1.000	0.917
	1	0.001	0.083	0.000	0.000	0.000	0.081
	2	0.000	0.018	0.000	0.000	0.000	0.002
	3	0.000	0.001	0.000	0.000	0.000	0.000
	4	0.000	0.000	0.000	0.000	0.000	0.000
$r = 1$	0	0.000	0.000	0.000	0.000	0.013	0.010
	1	0.999	0.810	1.000	1.000	0.986	0.931
	2	0.001	0.149	0.000	0.000	0.001	0.059
	3	0.000	0.029	0.000	0.000	0.000	0.000
	4	0.000	0.012	0.000	0.000	0.000	0.000
$r = 2$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.999	0.514	0.965	0.987	0.732	0.678
	3	0.001	0.318	0.031	0.013	0.268	0.322
	4	0.000	0.168	0.004	0.000	0.000	0.000
$r = 3$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.010	0.010
	2	0.000	0.001	0.236	0.275	0.227	0.302
	3	1.000	0.720	0.712	0.704	0.764	0.688
	4	0.000	0.279	0.053	0.021	0.000	0.000
$r = 4$	0	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.003	0.000	0.021	0.009	0.008	0.000
	3	0.100	0.007	0.053	0.050	0.062	0.086
	4	0.896	0.993	0.926	0.941	0.930	0.914

Note:
PIC: Posterior Information Criterion
KP^1: Kleibergen and Paap method with $\theta = 1.00$
KP^2: Kleibergen and Paap method with $\theta = 0.01$
KP^3: Kleibergen and Paap method with diffuse prior
BF^1: uses the Schwarz BIC given in (2.25)
BF^2: uses the harmonic mean of the likelihood given in (2.23)

Johansen: the numbers are p-values by Johansen's trace test

2.6 Illustrative Examples

In this section, we illustrate two examples of cointegration analysis using the method that is presented in previous sections. The main focus is to show the usefulness of our method with a relatively small number of observations and to compare it with other methods such as the PIC, the KP method and Johansen's test. The first example is a cointegration test for 'great ratios'. The second is for UK's PPP (purchasing power parity) and UIP (uncovered interest rate parity).

2.6.1 Cointegration Test for 'Great Ratios'

King *et al* (1991) (KPSW) examined cointegrating relationships between US output (Y), consumption (C), investment (I), and three other variables. In this sub-section, we investigate a three-variable model containing the real variables, C, I and Y. The data are quarterly and taken from the KPSW data set, which are: C (real per capita consumption, in logs), I (investment per capita, in logs), and Y (real private output per capita, in logs). We choose the shorter estimation period of 1968 (1) - 1988 (4), with a sample size of 83 to see how various tests choose the rank when the sample size is small. From economic theory, two cointegrating relations are expected to be found among these variables, given by $C - Y$ and $I - Y$, which are known as the 'great ratios'.

Table 2.5 presents the results of cointegration tests by the PIC, the KP method with $\theta = 1.00$ and with $\theta = 0.001$, Bayes factors calculated using

(2.25) and (2.23), and Johansen's trace test, with 2 lags[4] in VAR for the three-dimensional vector of time series $Y_t = \left[\begin{array}{ccc} C_t & I_t & Y_t \end{array} \right]$ with an intercept term. The prior specifications are the same as those used in the simulations. Thus, these prior hyperparameters favour no cointegration but are relatively noninformative given the fairly large variance. We assign an equal prior probability to each rank. We also impose restrictions on the cointegrating vector as $\beta = \left(\begin{array}{cc} I_r & \beta_\star \end{array} \right)$ for both identification and normalisation. From Table 2.5, most of the Bayesian tests show that the posterior probabilities for rank 1 are the highest for the PIC, KP^1 (with $\theta = 1.00$), KP^3 (with diffuse priors) and two BFs, while KP^2 (with $\theta = 0.001$) selects rank 0. There is almost no evidence of rank 2 or 3. The Bayesian tests find that there is a cointegration relationship between consumption and income, but not between investment and income, although KPSW found two cointegration relationships using full data set. On the classical side, Johansen's trace test cannot reject $r = 0$ at either 5 or 10 per cent significance level.

Table 2.6 shows the posterior results of $\beta_\star$ and α. If we assume the rank is 1, we expect that one cointegrating vector would be the first 'great ratio', which is the consumption-income relation, that is,

$$\beta' = \left(\begin{array}{ccc} 1 & \beta_{\star 1} & \beta_{\star 2} \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & -1 \end{array} \right) \tag{2.26}$$

The posterior means are close to these economic relations. Figure 2.1 presents the posterior densities of $\beta_\star$, which show that the expected cointegrating vector, $\beta_{\star 1} = 0$ and $\beta_{\star 2} = -1$, lies within the 95 per cent highest posterior

[4] Although KPSW select the VAR order as $p = 6$, the Schwarz Bayesian criterion selects the order 2 with the subset of their data.

Table 2.5: Cointegration Tests for the 'Great Ratios': Posterior Probabilities

rank r	PIC	KP^1	KP^2	KP^3	BF^1	BF^2	Johansen
0	-2.748	0.238	1.000	0.073	0.211	0.030	0.122
1	-2.881	0.694	0.000	0.914	0.762	0.968	0.651
2	-1.573	0.068	0.000	0.013	0.027	0.002	0.804
3	-1.011	0.000	0.000	0.000	0.000	0.000	–

Note:
PIC: Posterior Information Criterion
KP^1: Kleibergen and Paap method with $\theta = 1.00$
KP^2: Kleibergen and Paap method with $\theta = 0.01$
KP^3: Kleibergen and Paap method with diffuse prior
BF^1: uses the Schwarz BIC given in (2.25)
BF^2: uses the harmonic mean of the likelihood given in (2.23)

Johansen: the numbers are p-values by Johansen's trace test

Table 2.6: Bayesian Estimated $\beta_\star$ and α for the 'Great Ratios'

	$\beta_{\star 1}$	$\beta_{\star 2}$	α_1	α_2	α_3
Mean	-0.080	-0.947	0.135	0.327	0.213
s.d	0.144	0.190	0.031	0.057	0.040

density regions. Figure 2.2 presents the posterior densities of each element of α, which are skewed and lie far from zero. Figure 2.3 plots the cointegration relationship, which shows slightly upward trending.

The over-identifying restriction on the cointegrating vector is tested computing the Bayes factors using (2.25) and (2.23). The computed Bayes factors are 138.87 and 97.4 respectively. Therefore, there is strong evidence to support the consumption-income relation (see the guideline in Table 1.1, reproduced from Kass and Raftery, 1995).

Figure 2.1: Posterior Density of β for the 'Great Ratios'

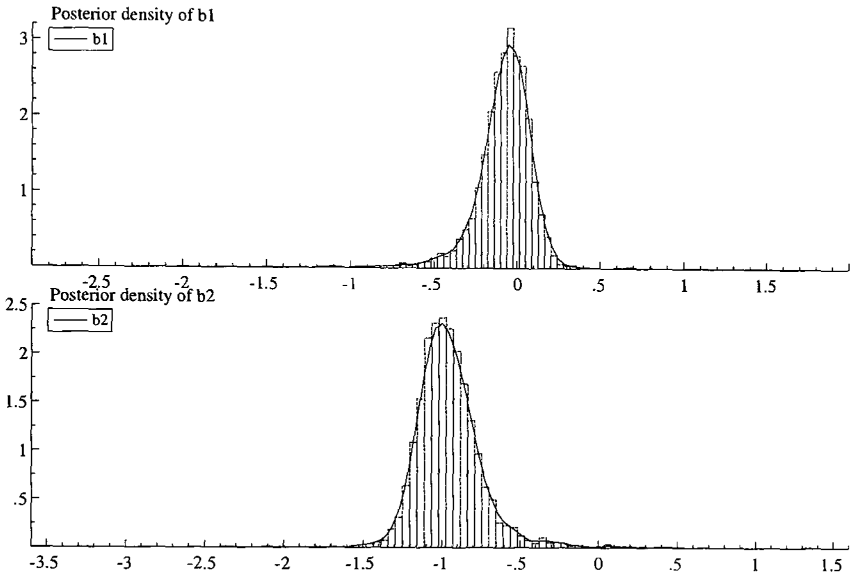


Figure 2.2: Posterior Density of α for the 'Great Ratios'

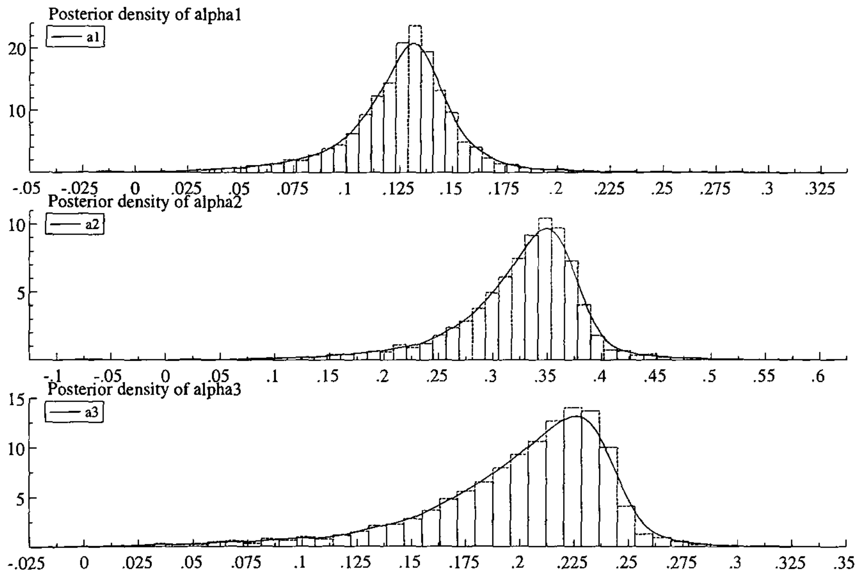
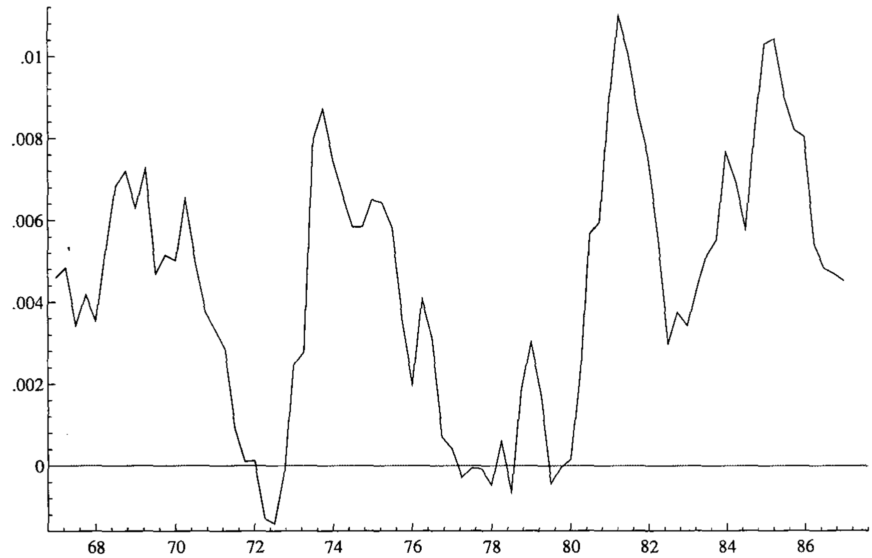


Figure 2.3: Cointegration Relationship for the 'Great Ratios'



2.6.2 Cointegration Test for PPP and UIP

Johansen and Juselius (JJ) (1992) studied cointegration for UK's PPP and UIP hypotheses by using likelihood ratio tests. In this subsection PPP and UIP hypotheses are tested by the various Bayesian method. The data are quarterly and have the following five variables: P (log of UK wholesale price), PF (log of trade weighted foreign wholesale price), R (three-month treasury bill rate in the UK), RF(three-month Eurodollar interest rate), E (log of UK effective exchange rate). As JJ conditioned their model on changes in oil prices and quarterly seasonal dummies, DPO (changes in real oil prices), DPO(-1) (changes in real oil prices with one lag), and $S1, S2, S3$ (quarterly seasonal dummies) are also included in the model as exogenous variables. The sample period is 1972(1) to 1987(2) with 62 observations. The VAR model for $Y_t = \begin{bmatrix} P_t & RF_t & PF_t & R_t & E_t \end{bmatrix}'$ is with 2 lags (chosen by SBIC) and an intercept term. Bayesian models are constructed with the same manner as the prior specifications in the previous example. Long-run economic theory suggests that two cointegrating vectors are expected as:

$$\beta' = \begin{bmatrix} I_2 & \beta_\star \end{bmatrix} = \begin{bmatrix} 1 & 0 & b1 & b2 & b3 \\ 0 & 1 & b4 & b5 & b6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \tag{2.27}$$

The first rows of (2.27) represents the PPP relation and the second is the UIP relation.

Table 2.7 shows results of the cointegration tests. Most methods select rank 2, except the KP^2 (with $\theta = 0.001$) which finds rank is 1. In the classical test, the trace test rejects the null of $r \leq 1$ against the alternative of $r \geq 2$ and cannot reject the null of $r \leq 2$ against $r \geq 3$ with 5 per cent significant level, suggesting $r = 2$.

Table 2.8 reports the posterior means and standard deviations for each unrestricted element of β and α estimated by Bayesian method proposed in previous section. Figure 2.4 and 2.5 plot the posterior densities of the cointegrating vector $\beta_\star$ and the adjustment term α. The first five densities of α (a1 - a5) correspond to the first cointegrating vector and the rest (a6 - a10) to the second cointegrating vector.

Next we consider testing for over-identifying restrictions on the cointegrating vectors. Table 2.9 reports the results of three Bayesian over-identifying restrictions tests. The first test is imposing the PPP relation (the first row in (2.27) on β to test whether the PPP relation holds. Bayes factors using (2.25) and (2.23) for this restrictions are 0.131 and 0.022 respectively, which shows no evidence for supporting the restrictions for the PPP relation. The second restriction is for testing the UIP relation by imposing the restrictions of the UIP relation on the cointegrating vectors. Bayes factors are 743.8 and 108.4, which shows very strong evidence to support the UIP relation according to the guideline in Table 2.1. The last testing for the over-identifying restrictions is the joint restrictions for the PPP and the UIP relations. Bayes factors are 0.001 and 0.0003, which does not support the expected cointegrating vectors in (2.27). Therefore, although the Bayesian test selects $r = 2$ as economic theory suggests, the PPP relation does not hold while the only UIP relation might hold.

Table 2.7: Cointegration Tests for PPP and UIP: Posterior Probabilities

rank r	PIC	KP[1]	KP[2]	KP[3]	BF[1]	BF[2]	Johansen
0	-7.183	0.000	0.066	0.088	0.101	0.013	0.004**
1	-0.737	0.033	0.831	0.385	0.111	0.094	0.034*
2	-0.755	0.580	0.103	0.513	0.788	0.893	0.058
3	-0.712	0.387	0.000	0.014	0.000	0.000	0.176
4	-0.693	0.000	0.000	0.000	0.000	0.000	0.023
5	-0.710	0.000	0.000	0.000	0.000	0.000	–

Note:
PIC: Posterior Information Criterion
KP[1]: Kleibergen and Paap method with $\theta = 1.00$
KP[2]: Kleibergen and Paap method with $\theta = 0.01$
KP[3]: Kleibergen and Paap method with diffuse prior
BF[1]: uses the Schwarz BIC given in (2.25)
BF[2]: uses the harmonic mean of the likelihood given in (2.23)

Johansen: the numbers are p-values by Johansen's trace test

Table 2.8: Posterior Results of $\beta_\star$ and α ($\lambda = 1.00$) for PPP and UIP

	Mean	s.d
$\beta_{\star 1}$	-1.139	0.053
$\beta_{\star 2}$	-3.712	1.191
$\beta_{\star 3}$	-0.545	0.118
$\beta_{\star 4}$	-0.074	0.010
$\beta_{\star 5}$	-0.590	0.323
$\beta_{\star 6}$	0.154	0.027

	Mean	s.d		Mean	s.d
α_1	-0.066	0.031	α_6	0.042	0.054
α_2	0.217	0.052	α_7	0.058	0.056
α_3	0.015	0.058	α_8	-0.005	0.115
α_4	-0.193	0.085	α_9	0.022	0.091
α_5	-0.022	0.028	α_{10}	-0.442	0.219

Figure 2.4: Posterior Densities of β for the PPP & UIP

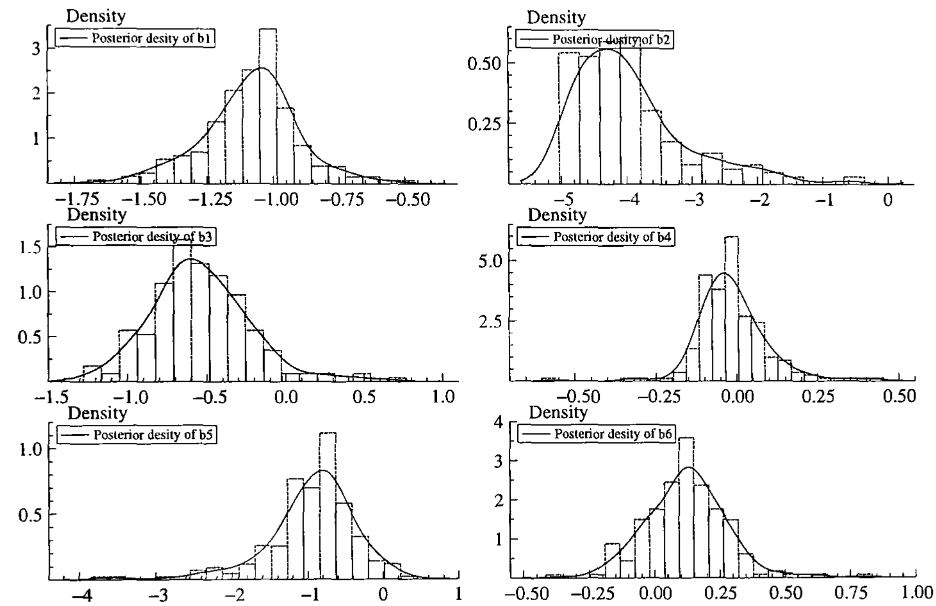


Figure 2.5: Posterior Densities of α for the PPP & UIP

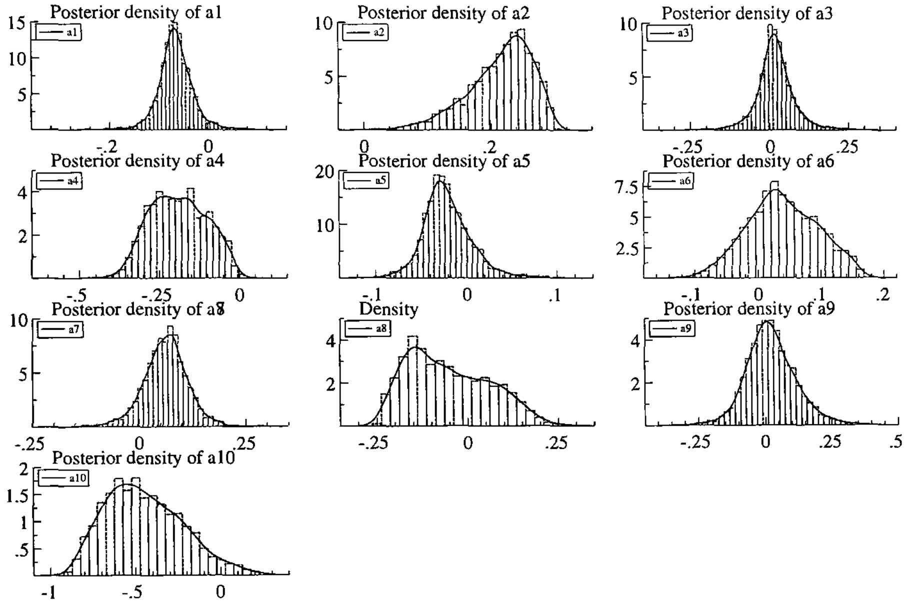


Figure 2.6: Cointegration Relationships for the PPP & UIP

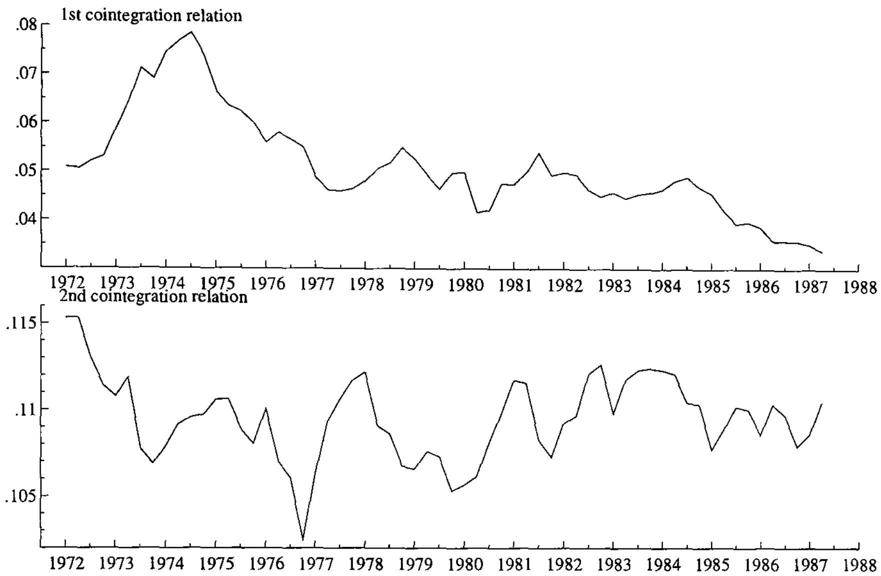


Table 2.9: Over-Identifying Restrictions on β: values of Bayes Factors for PPP and UIP

Restrictions	Test	BF^1	BF^2
$\beta' = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & * & * & * \end{bmatrix}$	PPP	0.131	0.022
$\beta' = \begin{bmatrix} 1 & 0 & * & * & * \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$	UIP	743.8	108.4
$\beta' = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$	PPP&UIP	0.001	0.0003

Note: "$\star$" in β' denotes the parameter to be estimated (unrestricted)

2.7 Conclusion

This chapter introduced a simple method of Bayesian cointegration analysis, and compares this method with other Bayesian methods such as the PIC and the KP method. The Bayes factors are used for computing the posterior probabilities for each rank using approximation method with Schwarz BIC and the harmonic mean of the likelihood, which provide insensitive in the choice of the prior parameters. Monte Carlo simulations show that the Bayes factors computed by using the harmonic mean of the likelihood tend to select the correct cointegrating rank than those by using the Schwarz BIC approximation method when the sample size is small. However, when the sample size is large, the Bayes factors approximated by the SBC performs slightly better. The Bayes factors are also applied to test over-identifying restrictions on the cointegrating vectors.

For the comparison with the PIC, we find that the performance of the PIC depends upon the sample size as we expected. In the Monte Carlo experiments, the PIC shows better performance when the true rank is 0 or 1, however, the performance is much worse when the true rank is higher. With large sample size, the PIC shows the best performance among others.

The posterior odds by the KP method with conjugate priors relies on the specification of the prior hyperparameters. Although we did not conduct the prior sensitivity test formally, the Monte Carlo simulations reveal that choices of the prior parameters largely affect the values of Bayes factors. However, with diffuse priors, the KP shows good performance comparable to that of the PIC.

For the future research, it is of interest that we consider Strachan (2003)'s or Villani(2003)'s method with valid prior on the cointegration vectors.

Chapter 3

Markov Switching Cointegration Model

3.1 Introduction

The previous chapter deals with a linear cointegration model with a Bayesian approach. This chapter introduces a Markov switching cointegration model that allows the cointegration relationships to be switched on and off depending on the regime, and its application to purchasing power parity (PPP). We also present a cointegration model in which deviations from the long-run equilibrium are characterised different rates of adjustment depending on the regime. Many economic theories are concerned with equilibrium relationships in which several series are expected to be cointegrated each other. However, it is sometimes not possible to find such cointegrating relationships because of presence of transaction costs, adjustment costs, or government's policy change. Especially in the goods market disequilibria may take a considerable amount of time to be reversed (Brenner and Kroner, 1995). There are several statistical explanations for failing to reject the null of no cointegration due to the span of the data set (Hendry, 1995), structural breaks (Gregory and Hansen, 1996, and Campos, *et al.,* 1996), and the choice of the number of lags in the VAR (Banerjee, *et al.*, 1993).

Recently, several authors have proposed cointegration models where cointegration occurs temporarily. The concept of the temporal cointegration was proposed by Granger and Siklos (1996). The authors show that cointegration can be switched off because a common stochastic trend is added to the model, and find that cointegration of the US-Canada interest rate parity does not occur globally but it occurs only under the regime of inflation targeting of Canada. The authors select *ex ante* breaking point that divides the given series into two groups, then conduct Johansen's LR test for each regime. This method requires a priori information about the location of the shifts in regime. Thus, if exact dates of shift are not available, their method cannot be applied. Furthermore, if there are several shifts, it incurs power problem of Johansen's test since smaller samples are available for each regime. Thus, if those problems arise, the researcher must resort to other methods. Balke and Fomby (1997) considered threshold cointegration in order to investigate a model in which there is discontinuous adjustment to a long-run equilibrium. They chose a threshold model based on the idea that only when the deviation from the equilibrium exceeds a critical threshold, do the benefits of adjustment exceed the costs and, hence economic agents act to move the system back toward the equilibrium. To investigate the model, they suggest testing for cointegration first and then examining the error correction model for signs of Markov switching behaviour. That is, after finding the cointegration relation β based on the linear cointegration model, set up the cointegration relationships $\beta' X_t = z_t$ and then investigate whether $\{z_t\}$ follows a threshold process with one stationary, and one non-stationary regime. This two-step method[1] is asymptotically valid but might yield unreliable estimation for cointegrating vectors when the sample size is small and/or when a regime where cointegration is present is not dominant over the sample period. Suppose we analyse Markov switching VECM such as $\triangle y_t = \alpha_{s_t}\beta'_{NL}y_{t-1} + e_t$ where the state variable s_t takes 1 if cointegration is present and 0 if not; $\alpha_{s_t=1} = \alpha_1 \neq \mathbf{0}$ and $\alpha_{s_t=0} = \mathbf{0}$. The cointegrating vector, β_{NL}, might be exactly estimated by modeling linear cointegrated VECM $\triangle y_t = \alpha\beta'_L y_{t-1} + e_t$ if enough large observations for a regime when cointegration is present are available in the Markov switching cointegration model. However, if only a small sample size is available, estimated $\widehat{\beta}_L$ will be largely biased from the true β_{NL}. Thus, a classical method such as Balke and Fomby could be misleading. There might be a method that can estimate β_{NL} directly within the nonlinear models in classical framework such as using a grid search method. However, it will not be easy. To overcome this problem, we employ a Bayesian method with Markov Chain Monte Carlo (MCMC) simulation techniques to estimate parameters such as β_{NL} conditional on the estimated set of regime variables $S = \{s_1, s_2, \ldots, s_T\}'$ in each iteration of the Gibbs sampler so that we can obtain more accurate estimated values of β_{NL}.

In Markov switching models, some parameters are dependent on an un-

[1]Psaradakis, *et al* (2003) use this two-step method to analysis a two-regime Markov error-correction model that allows for different rates of adjustment to long-run equilibrium.

observed regime variable s_t that is an outcome of discrete Markov process. In the classical methods, inference on the unobserved Markov switching variable $S = \{s_1, s_2, \ldots, s_T\}'$ is made on the conditional distribution after making inferences on the model's unknown parameters. In Bayesian method, both the parameters of the model and the regime variables S are treated as random variables, and thus inference on S is based on a joint distribution.

Applying a Bayesian approach has another advantage for a nonlinear cointegration model. Bayes factors enable to test for a no-cointegrated linear model against Markov switching cointegration models, while in classical methods it is difficult to test because of nonstandard inference problem due to the presence of unit roots and the unidentifiability of the nuisance parameters under the null hypothesis. By taking this advantage of the Bayes factors, we can select the most appropriate model among more various models under consideration.

The plan of the paper is as follows. Section 3.2 presents the statistical model for the Markov switching cointegration and estimation method. We specify the prior densities, likelihood functions, and then derive the posterior distributions. Section 3.3 describes testing for Markov switching and model selection by Bayes factors. Then we show simulated experiments with artificially generated data to evaluate how the proposed method can detect an appropriate model and estimate the cointegrating vector correctly. Section 3.4 illustrates applications to a simple PPP model between UK-USA. This example reveals the problem of estimation for β_{NL} if we employ the classical method such as Balke and Fomby. It shows that with relatively small number of observations the estimated β_L of the linear cointegration model is significantly different from the estimated β_{NL} by the Bayesian Markov switching cointegration model with the posterior distribution for β_{NL} conditional on the regime variables. Section 3.5 contains concluding remarks and suggestions for future work.

3.2 Bayesian Estimation of the Markov Switching Cointegration Model

3.2.1 Markov Switching Cointegration Model with Two-Regime

This section proposes a Markov switching cointegration model. Let X_t denote an $I(1)$ vector of n-dimensional time series with r linear cointegrating relations. The long-run multiplier matrix is $\Pi = 0_n$ when the regime variable, s_t, takes its value zero, and $\Pi \neq 0_n$ when $s_t = 1$. If we assume that the intercept term μ in Gaussian VAR is also subject to Markov switching, then the VECM representation is:

$$\Delta X_t = \mu_{s_t} + \Pi_{s_t} X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{3.1}$$

where $t = p, p+1, \ldots, T$, and p is the number of lags, and the errors ε_t are assumed $N(0, \Sigma)$ and independent over time. Dimensions of matrices are μ and ε $(n \times 1)$, Π, Ψ and Σ $(n \times n)$. The state variable s_t evolves according to a two-state, first-order Markov switching process with the transition probabilities, $p\left(s_t = 0 \mid s_{t-1} = 0\right) = p_{00}$ and $p\left(s_t = 1 \mid s_{t-1} = 1\right) = p_{11}$.

More specifically, equation (3.1) is:

$$\Delta X_t = \mu_0 + \mu_1 s_t + \alpha\beta' \left(I_n \otimes s_t\right) X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{3.2}$$

where α and β are $n \times r$. Under the regime 0 when there is no cointegration, the intercept parameter takes the value μ_0 and under the second regime the intercept term takes the value $\mu_0 + \mu_1$.

Equation (3.2) can be rewritten in the matrix format as:

$$Y = X\Gamma + Z\beta\alpha' + E = WB + E \tag{3.3}$$

where

$$Y = \begin{bmatrix} \Delta X_p' \\ \Delta X_{p+1}' \\ \vdots \\ \Delta X_T' \end{bmatrix}, Z = \begin{bmatrix} X_{p-1}' \left(I_n \otimes s_p\right) \\ X_p' \left(I_n \otimes s_{p+1}\right) \\ \vdots \\ X_{T-1}' \left(I_n \otimes s_T\right) \end{bmatrix}, E = \begin{bmatrix} \varepsilon_p' \\ \varepsilon_{p+1}' \\ \vdots \\ \varepsilon_T' \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} \mu_0' \\ \mu_1' \\ \Psi_1' \\ \vdots \\ \Psi_{p-1}' \end{bmatrix}, X = \begin{bmatrix} 1 & s_p & \Delta X_{p-1}' & \cdots & \Delta X_1' \\ 1 & s_{p+1} & \Delta X_p' & \cdots & \Delta X_2' \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & s_T & \Delta X_{T-1}' & \cdots & \Delta X_{T-p+1}' \end{bmatrix},$$

$$W = \begin{bmatrix} X & Z\beta \end{bmatrix}, B = \begin{bmatrix} \Gamma \\ \alpha' \end{bmatrix}.$$

Let τ be the number of rows of Y, so that $\tau = T - p + 1$, then X is $\tau \times (2 + n(p-1))$, Γ is $(2 + n(p-1)) \times n$, W is $\tau \times k$ where $k = 2 + n(p-1) + r$, and B is $k \times n$. Equation (3.3) represents the multivariate regression format of (3.1).

3.2.2 Markov Switching Cointegration Model with m-Regime

We consider in this chapter a Markov switching cointegration model as in (3.1), however, this model is restrictive as imposing zeros in the long-run multiplier matrix in one regime. Rather than imposing the restriction of no cointegration in one regime, we can consider a more general model in which deviations from the long-run equilibrium are characterised by different rates of adjustment depending upon the regime. The modification for this general model from a model (3.1) is easy as follows.

The long-run multiplier matrix is decomposed as $\alpha_{s_t}\beta'$, both are $n \times r$, where α_{s_t} is the adjustment term that is subject to the regime $s_t = i$ where $i = 1, 2, \ldots, m$ and β' is the cointegrating vector. If we assume that the intercept term μ and trend ξ in VAR are also subject to Markov switching, then the VECM representation is:

$$\Delta X_t = \mu_{s_t} + \xi_{s_t} t + \alpha_{s_t}\beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{3.4}$$

where $t = p, p+1, \ldots, T$, and p is the number of lags, and ε_t are assumed $N(0, I_n)$ and independent over time. Dimensions of matrices are μ and ε $(n \times 1)$, Ψ and Σ $(n \times n)$. The state variable s_t evolves according to a m-state, first-order Markov switching process with the transition probabilities, $p\left(s_t = i \mid s_{t-1} = j\right) = p_{ij}$.

Equation (3.4) can be rewritten in the matrix format as:

$$Y = X\Gamma + Z\beta\alpha'_{s_t} + E = WB + E \tag{3.5}$$

where

$$Y = \begin{bmatrix} \Delta X'_p \\ \Delta X'_{p+1} \\ \vdots \\ \Delta X'_T \end{bmatrix}, Z = \begin{bmatrix} X'_{p-1} \\ X'_p \\ \vdots \\ X'_{T-1} \end{bmatrix}, E = \begin{bmatrix} \varepsilon'_p \\ \varepsilon'_{p+1} \\ \vdots \\ \varepsilon'_T \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} \mu_1 & \dots & \mu_m & \xi_1 & \dots & \xi_m & \Psi_1 & \dots & \Psi_{p-1} \end{bmatrix},$$

$$X = \begin{bmatrix} s_{1,p} & \cdots & s_{m,p} & s_{1,p} & \cdots & s_{m,p} & \Delta X'_{p-1} & \cdots & \Delta X'_1 \\ s_{1,p+1} & \cdots & s_{m,p+1} & 2s_{1,p+1} & \cdots & 2s_{m,p+1} & \Delta X'_p & \cdots & \Delta X'_2 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{1,T} & \cdots & s_{m,T} & (T-p+1)s_{1,T} & \cdots & (T-p+1)s_{m,T} & \Delta X'_{T-1} & \cdots & \Delta X'_{T-p+1} \end{bmatrix}$$

$$W = \begin{bmatrix} X & \mathcal{I}_1 Z\beta & \mathcal{I}_2 Z\beta & \cdots & \mathcal{I}_m Z\beta \end{bmatrix},$$

$$B = \begin{bmatrix} \Gamma' & \alpha_1 & \alpha_2 & \dots & \alpha_m \end{bmatrix}'$$

Let τ be the number of rows of Y, so that $\tau = T - p + 1$, then X is $\tau \times (m + n(p-1))$, Γ is $(m + n(p-1)) \times n$, W is $\tau \times k$ where $k = m + n(p-1) + mr$, and B is $k \times n$. $s_{i,j}$ in X is an indicate variable such that it equals to 1 if regime is i and 0 otherwise. $\mathcal{I}_i$ in W is an indicator matrix $(\tau \times \tau)$ where the diagonal elements are 1 if t^{th} regime is i, otherwise 0 and the rest of the elements are 0. m is the number of regimes.

3.2.3 Prior Distributions and Likelihood Functions

From this subsection and following subsection, we present how we estimate Markov switching cointegration model with two-regime of (3.2) or more general m-regime model of (3.4). First of all, prior distributions for all parameters should be specified. We choose a Student-t prior for β, diffuse (non-informative) prior for regime variable $S = \{s_1, s_2, \ldots, s_T\}'$, and the natural conjugate priors for B, thus:

$$\Sigma \sim IW\left(\eta, h\right) \tag{3.6}$$

$$B \mid \Sigma \sim MN\left(P, \Sigma \otimes A^{-1}\right) \tag{3.7}$$

$$\beta_{\star} \sim Mt\left(\overline{\beta}, H, Q, v\right) \tag{3.8}$$

$$p_{00} \sim beta\left(u_{00}, u_{01}\right) \tag{3.9}$$

$$p_{11} \sim beta\left(u_{11}, u_{10}\right) \tag{3.10}$$

where IW refers to an inverted Wishart distribution with parameters $\eta \in \mathbb{R}^{n\times n}$ and degrees of freedom, v; MN refers to a matrix normal with mean $P \in \mathbb{R}^{k\times n}$, $k = n(p-1)+r+1$, $A \in \mathbb{R}^{k\times k}$; Mt refers to a matrix Student-t distribution with parameters $\overline{\beta} \in \mathbb{R}^{(n-r)\times r}$, $Q \in \mathbb{R}^{n-r}$, $H \in \mathbb{R}^{r}$; *beta* refers to a beta distribution. Note that r^2 restrictions for identification and normalisation are imposed on β such that $\beta' = (\ I_r \quad \beta_\star'\ )$, where $\beta_\star$ is $(n-r) \times r$ unrestricted matrix as in the previous chapter.

Instead of above priors, if we do not want to impose an informative prior for Σ, we can opt diffuse prior for Σ as $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$. Also, we can specify the priors for transition probabilities, p_{00} and p_{11} as $p(p_{00}) \propto 1_{(0,1)}$, $p(p_{11}) \propto 1_{(0,1)}$, where $1_{(0,1)}$ represents an indicator function that is 1 on the interval (0, 1) and 0 otherwise.

The likelihood function for B, Σ, S and β is given by,

$$\begin{aligned}\mathfrak{L}\left(S, B, \Sigma, \beta \mid Y\right) &\propto |\Sigma \otimes I_t^{-1}|^{-t/2} \exp\left[-\frac{1}{2}\left\{vec\left(Y-WB\right)'\left(\Sigma^{-1}\otimes I_t^{-1}\right) vec\left(Y-WB\right)\right\}\right] \\ &\propto |\Sigma|^{-t/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left\{\Upsilon + (B-\widehat{B})'W'W(B-\widehat{B})\right\}\right]\right\}\end{aligned} \tag{3.11}$$

where $\widehat{B} = (W'W)^{-1}W'Y$, and $\Upsilon = (Y - W\widehat{B})'(Y - W\widehat{B})$.

The likelihood function for the transition probabilities p_{00}, p_{11}, which are independent of the data but conditional on the set of the regime variables, is given by:

$$\mathfrak{L}\left(p_{00}, p_{11} \mid S\right) = p_{00}^{m_{00}}\left(1-p_{00}\right)^{m_{01}} p_{11}^{m_{11}}\left(1-p_{11}\right)^{m_{10}} \tag{3.12}$$

where $m_{i,j}$, $i, j = 0$ or 1, denotes the number of the transition from the regime i to j, that can be counted from given S.

3.2.4 Posterior Specifications

In this subsection we derive the posterior density from the priors and the likelihood functions. First, we show generating the regime variable S by the multi-move Gibbs sampler, then derive the posterior distributions for other parameters conditional on S. The other method to generate draws of the regimes is by Albert and Chib (1993), and Paap and van Dijk (2003) chose this algorithm for a Markov trend cointegration model.

To sample the regime variable S we employ the multi-move Gibbs sampling method, which was originally proposed by Carter and Kohn (1994) and applied to univariate Markov switching models by Kim and Nelson (1998). The multi-move Gibbs sampling refers to simulating s_t, $t = p, p+1, \ldots, T$, as a block from the following conditional distribution:

$$p(S \mid \Theta, Y) = p(s_T \mid \Theta, Y) \prod_{t=p}^{T-1} p(s_t \mid s_{t+1}, \Theta, Y) \tag{3.13}$$

where $\Theta = \{B, \Sigma, \beta, p_{00}, p_{11}\}$. The first term of the right hand side of the above equation, $p(s_T \mid \Theta, Y)$, can be obtained from running the Hamilton filter (Hamilton, 1989). To draw s_t conditional on s_{t+1}, Θ and Y, we use the following results:

$$p(s_t \mid s_{t+1}, \Theta, Y) = \frac{p(s_{t+1} \mid s_t, \Theta, Y)\, p(s_t \mid \Theta, Y)}{p(s_{t+1} \mid \Theta, Y)} \propto p(s_{t+1} \mid s_t)\, p(s_t \mid \Theta, Y) \tag{3.14}$$

where $p(s_{t+1} \mid s_t)$ is the transition probability, and $p(s_t \mid \Theta, Y)$ is taken from the Hamilton filter. Using Equation (3.14) we compute:

$$Pr\left(s_t = 1 \mid s_{t+1}, \Theta, Y\right) = \frac{p\left(s_{t+1} \mid s_t = 1\right) p\left(s_t = 1 \mid \Theta, Y\right)}{\sum_{j=0}^{1} p\left(s_{t+1} \mid s_t = j\right) p\left(s_t = j \mid \Theta, Y\right)} \tag{3.15}$$

Once above probabilities are computed, we draw a random number from a uniform distribution between 0 and 1, and if the generated number is less than or equal to the value calculated by (3.15), we set $s_t = 1$, otherwise, set equal to 0.

After drawing S by multi-move Gibbs sampling, we generate the transition probabilities, p_{00} and p_{11} by multiplying (3.9) and (3.10) by the likelihood function (3.12)

$$p\left(p_{00}, p_{11} \mid S\right) \propto p_{11}^{u_{11}+m_{11}-1}\left(1 - p_{11}\right)^{u_{10}+m_{10}-1} p_{00}^{u_{00}+m_{00}-1}\left(1 - p_{00}^{u_{01}+m_{01}-1}\right) \tag{3.16}$$

The posterior of the transition probabilities in (3.16) assumes that, conditional on S, the transition probabilities, p_{00} and p_{11}, are independent of both the other parameters of the model and the data, Y.

Next, we consider the posteriors of B, Σ and β. The joint prior of B, Σ and $\beta_\star$ is given by multiplication of (3.6), (3.7) and (3.8) as follows:

$$p(B, \Sigma, \beta_\star) \propto g\left(\beta_\star\right)|A|^{n/2}|\Sigma|^{-\frac{k+h+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[\eta + (B-P)'A(B-P)\right]\right\}\right], \tag{3.17}$$

where $g(\beta_\star)$ denotes the prior for the cointegrating vector.

Next, we consider conditional posterior densities for Σ, B and β. By multiplying the joint prior in (3.17) by (3.11) conditional on regime variable S, we have

$$\begin{aligned}
&p(B, \Sigma, \beta_\star \mid S, Y) \\
&\propto g(\beta_\star)\,|\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[\eta + \Upsilon + (P-\widehat{B})'[A^{-1} + (W'W)^{-1}]^{-1}(P-\widehat{B})\right.\right.\right. \\
&+(B-B_\star)'A_\star(B-B_\star)]\}] \\
&= g(\beta_\star)\,|\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[\Xi + (B-B_\star)'A_\star(B-B_\star)\right]\right\}\right] \qquad (3.18)
\end{aligned}$$

where $c = t+k+n+1$, $A_\star = A + W'W$, $B_\star = (A+W'W)^{-1}(AP + W'W\widehat{B})$, and $\Xi = \eta + \Upsilon + (P-\widehat{B})'[A^{-1} + (W'W)^{-1}]^{-1}(P-\widehat{B})$.

From (3.18), the conditional posterior of Σ is derived as an inverted Wishart distribution, and the conditional posterior of B as a matrix-variate normal density with covariance, $\Sigma \otimes A_\star^{-1}$, that is,

$$p(\Sigma \mid \beta_\star, S, Y) \propto |\Xi|^{t/2}|\Sigma|^{-(t+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\Xi\right)\right] \qquad (3.19)$$

$$p(B \mid \Sigma, \beta_\star, S, Y) \propto |A_\star|^{n/2}|\Sigma|^{-k/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(B-B_\star)'A_\star(B-B_\star)\right\}\right] \qquad (3.20)$$

Thus, the conditional posterior distributions for Σ and B are given as (3.19) and (3.20) respectively.

The joint posterior of B and $\beta_\star$ can be derived by integrating (3.18) with respect to Σ,

$$\begin{aligned} p(B, \beta_\star \mid S, Y) &\propto \int p(B, \Sigma, \beta_\star \mid Y) d\Sigma \\ &\propto \int g\left(\beta_\star\right) |\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[\Xi + (B - B_\star)' A_\star (B - B_\star)\right]\right\}\right] d\Sigma \\ &\propto g\left(\beta_\star\right) |\Xi + (B - B_\star)' A_\star (B - B_\star)\ |^{-(t+k+1)/2} \end{aligned} \tag{3.21}$$

By integrating (3.21) with respect to B we obtain the posterior density of the cointegrating vector $\beta_\star$,

$$\begin{aligned} p(\beta_\star \mid S, Y) &\propto \int p(B, \beta_\star \mid Y) dB \\ &\propto \int g\left(\beta_\star\right) \mid \Xi + (B - B_\star)' A_\star (B - B_\star) \mid^{-\frac{t+k+1}{2}} dB \\ &\propto g\left(\beta_\star\right) \mid \Xi \mid^{-(t+1)/2} \mid A_\star \mid^{-n/2} \end{aligned} \tag{3.22}$$

To summarise, we have the following posterior distributions with conditions of S:

$$\Sigma \mid \beta, S, Y \sim IW\left(\Xi, t + v\right) \tag{3.23}$$

$$B \mid \Sigma, \beta, S, Y \sim MN\left(B_\star, \Sigma \otimes A_\star^{-1}\right) \tag{3.24}$$

$$p(\beta_\star \mid S, Y) \propto g(\beta_\star) \mid \Xi \mid^{-(t+h+1)/2} \mid A_\star \mid^{-n/2} \tag{3.25}$$

Conditional on the regime variable S, the Markov switching cointegration model in (3.2) or (3.4) are simply regression models with a known dummy variable, S.

The posterior distribution for $\beta_\star$ in (3.25) is not a known form and can be drawn by employing importance sampling, the Metropolis-Hastings algorithm (see Chib and Greenberg, 1995) or the griddy-Gibbs sampling (see Ritter and Tanner, 1992). In this chapter, we chose the griddy-Gibbs sampling technique as in the previous chapter because the algorithm does not require the specification of some function that approximate the distribution.

Next, we describe a Bayesian procedure for estimating the Markov switching cointegration model that is introduced in the previous section. A Gibbs sampling technique is employed for estimation of the model. Given the conditional posterior distributions, we implement the Gibbs sampling to generate sample draws. The following steps can be replicated.

- Step 1: Set $i = 1$. Specify starting values for the parameters of the model, $S^{(0)} = \left\{s_1^{(0)}, s_2^{(0)}, \ldots, s_T^{(0)}\right\}'$, $\beta^{(0)}$ and $\Sigma^{(0)}$, where Σ is a covariance matrix.

- Step 2: Generate $\beta_\star^{(i)}$, the unrestricted elements of β, from $p(\beta_\star \mid S^{(i-1)}, Y)$ using the Griddy Gibbs sampling algorithm (or Metropolis-Hastings/Importance sampling algorithm).

- Step 3: Generate $B^{(i)}$ from $p(B \mid \beta_\star^{(i)}, \Sigma^{(i-1)}, S^{(i-1)}, Y)$.

- Step 4: Generate $\Sigma^{(i)}$ from $p(\Sigma \mid \beta_{\star}^{(i)}, S^{(i-1)}, Y)$.

- Step 5: Generate $(p_{11}, p_{00})^{(i)}$ from $p\left(p_{00}, p_{11} \mid S^{(i-1)}\right)$, where p_{11} and p_{00} are the transition probabilities defined as $p_{11} = \Pr\left(s_t = 1 \mid s_{t-1} = 1\right)$ and $p_{00} = \Pr\left(s_t = 0 \mid s_{t-1} = 0\right)$

- Step 6: Generate $S^{(i)} = \left\{s_1^{(i)}, s_2^{(i)}, \ldots, s_T^{(i)}\right\}'$ from $p\left(S \mid \Theta^{(i)}, Y\right)$, where $\Theta = \{B, \Sigma, \beta_{\star}, p_{00}, p_{11}\}$, using multi-move Gibbs sampling algorithm.

- Step 7: Set $i = i + 1$, and go to Step 2.

Step 2 through Step 7 can be iterated N times to obtain the posterior densities. Note that the first L times iterations are discarded in order to attenuate the effect of the initial values.

3.3 Testing for Markov Switching and Model Selection by Bayes Factors

This section examines how one may detect Markov cointegration. In classical methods, testing for Markov switching in cointegrated models has several difficulties. First, as Balke and Fomby (1997) noted, for testing the no cointegration linear model against a nonlinear cointegration model, there is a nonstandard inference problem due to the presence of unit roots and the unidentifiability of the nuisance parameters (threshold values for threshold cointegration and the transition probabilities for Markov switching cointegration). As a result, the testing procedure suggested by them consists of two steps - first testing for no cointegration/cointegration, then second, testing for nonlinearity if the test rejects the null of no cointegration. This procedure excludes the case when one cannot find cointegration globally but can detect a Markov cointegration locally as in Granger and Siklos (1996). As Balke and Fomby showed, for larger threshold values or smaller sample sizes, both the augmented Dickey-Fuller and the Phillips-Perron tests have serious power problem. Thus, it is important to test for no cointegration linear model against a nonlinear cointegration model. As to testing linear cointegration against a Markov switching cointegration, one may use Hansen (1992 and 1996) or Garcia (1998) in order to overcome the non standard problem.

In Bayesian approach, these problems do not arise since Bayes factors integrate out nuisance parameters. Moreover, using Bayes factors has some advantages over classical methods such as testing for all the models under consideration or dealing with likelihoods with multiple peaks, as described in Koop and Potter (1998). With recent development of algorithms for computing Bayes factors, it has been more popular to use the Bayes factors (and accordingly, the posterior probabilities) for model selection and testing within Bayesian framework. Kass and Raftery (1995) provides a survey of Bayes factor. For the Bayesian approach to the Markov switching model, Koop and Potter (1998), Chib (1998) and Kim and Nelson (1999) used Bayes factors for testing nonlinearity and model selection in univariate models. Koop and Potter (1998) chose the Savage-Dickey density ratio (Dickey, 1971) to compute the marginal likelihood as the linear model is nested within the Markov switching or threshold model. Chib (1998) employed Chib's (1995) approach to calculating the marginal likelihoods that utilises the output from the Gibbs sampling. Kim and Nelson (1999) employ a method that is based on the sensitivity of the posterior probability of the model indicator parameter to the prior probability.

In this chapter we deal with testing for Markov switching in a multivariate model as a problem of model selection and approximate the Bayes factors by Schwarz's Bayes information criterion (BIC) to select the most appropriate model. The Schwarz BIC can give a rough approximation to the Bayes factors, which is easy to use and does not require evaluation of the prior distribution, as Kass and Raftery (1995) noted. Wang and Zivot (2000) employed the Schwarz BIC to compute the Bayes factors for detecting the number of structural breaks. The Schwarz BIC for model j can be computed as

$$\mathrm{BIC}_j = -2 \ln \mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right) + q_j \ln(t) \tag{3.26}$$

where $\mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right)$ denotes the likelihood function under the model j; q_j denotes the total number of estimated parameters in the model j; M_j denotes the model indicator for model j. The likelihood function $\mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right)$ is evaluated at $\widehat{\theta_j}$, the posterior means of the parameters for model j. The Bayes factor for model k against model j can be approximated by (2.25).

By using the Schwarz BIC to approximate to logarithm of the Bayes factor, it is easy to test Markov switching cointegration as a problem of model selection. In our case, we compute the Schwarz BIC conditional on the regime variable S such that

$$\begin{aligned}\mathrm{BIC}_j &= -2\ln\mathfrak{L}\left(S,B,\Sigma,\beta_\star,p_{00},p_{11}\mid Y;\mathrm{M}_j\right)+q_j\ln\left(t\right)\\ &= -2\left\{\ln\mathfrak{L}\left(S,B,\Sigma,\beta_\star\mid Y;\mathrm{M}_j\right)+\ln\mathfrak{L}\left(p_{00},p_{11}\mid S;\mathrm{M}_j\right)\right\}+q_j\ln\left(t\right)\end{aligned} \tag{3.27}$$

Within the framework of the Markov switching cointegration model, we test the Markov switching by modelling as following:

M0: Linear No-Cointegration Model.

M1: Linear Cointegration Model.

M2: Markov Cointegration Model with a constant intercept.

M3: Markov Cointegration Model with a regime dependent intercept

It is possible to consider more general models such as a heteroskedastic Markov cointegration/no cointegration model or homo/heteroskedastic Markov switching without cointegration. However, we focus on the above four models for simplicity. Note that linear models M0 and M1 can be thought as Markov switching cointegration model with restrictions on the transition probabilities such as $p_{00} = 1$ and $p_{11} = 0$ for M0 and $p_{00} = 0$ and $p_{11} = 1$ for M1.

It might be possible to compute Bayes factors for all model M0 - M3 to select the most appropriate model. However, if the true model is M0 or M1, computation of the Bayes factors for M2 - M3 might not be feasible because of the problem that some variables are not identified through the Gibbs sampling. For example, if generated regime variables are $s_t = 0$ for all t, then μ_1 and β cannot be identified. Kim and Nelson (1999) deal with this problem by employing 'pseudo priors' (see Carlin and Chib, 1995). In this paper we restrict a priori that a certain percentage of the observations lie in each regime in order to avoid the problem. When the number of one regime in the generated state variables $S^{(i)} = \left\{s_1^{(i)}, s_2^{(i)}, \ldots, s_T^{(i)}\right\}'$ in an iteration of the MCMC is less than given percentage, then the previously drawn values, $S^{(i-1)}$, is used again in the next iteration.

To evaluate how the proposed method can detect the correct model among M0 - M3, we conducted a Monte Carlo simulations. In our experiments, we consider three-variable Markov switching VECM. The data generating processes (DGPs) are given as the following:

$$\begin{aligned}
\text{M0}: &\quad \Delta X_t = \mu_1 + \varepsilon_t \\
\text{M1}: &\quad \Delta X_t = \mu_1 + \alpha_1 \beta' X_{t-1} + \varepsilon_t \\
\text{M2}: &\quad \Delta X_t = \mu_1 + \alpha_{s_t} \beta' X_{t-1} + \varepsilon_t \\
\text{M3}: &\quad \Delta X_t = \mu_{s_t} + \alpha_{s_t} \beta' X_{t-1} + \varepsilon_t
\end{aligned}$$

where $s_t = 1$ if cointegration is present and $s_t = 0$ if not; μ_1=(0.2, 0.2, 0.2)', μ_0=(-0.2, -0.2, -0.2)', α_1=(-0.2, -0.2, -0.2)', α_0=(0, 0, 0)', β'=(1, -1, 1), $\varepsilon_t \sim NID(0,\, I_3)$, and the sample size T={100, 200, 500}. The transition probabilities are (p_{11}, p_{00})∈{(0, 1), (1, 0), (0.90, 0.98), (0.95, 0.90), (0.95, 0.95)}. The first pair of transition probabilities, (0, 1), is when the true model is M0. The second pair of transition probabilities, (1, 0), is when the true model is M1. The third pair of transition probabilities, (0.90, 0.98) implies that the regime 0 when cointegration is not present is much longer than the regime 1 when cointegration is present, thus it can be considered as testing for no-cointegration linear model against Markov switching cointegration. On the other hand, the fourth pair, (0.95, 0.90), implies the regime 1 is dominant over the regime 0, and thus it can be considered as testing for linear cointegration model against Markov switching cointegration. The fifth pair, (0.95, 0.95), will produce symmetric persistence of the two regimes.

The Bayes factors were computed for all models (M0 - M3) to calculate the posterior probability for each model. For prior parameters, we set $\eta = I_3/1000$ for covariance prior in (3.6), $A = I_k/1000$ in (3.7) and $H = 1/1000$ in (3.8) to ensure fairly large variance for representing prior ignorance, $\overline{\beta} = 0_{2\times1}$, $v = 3.001$, $u_{00} = u_{11} = 9$, $u_{01} = u_{10} = 1$, $Q = I_3$, $P = 0$ favouring the absence of cointegration. We have used different sets of prior parameters for η, A and H and found no significant effects on the conclusions due to the SBC's insensitivities of prior choices. The number of cointegration rank and the number of the lags in VAR is assumed to be known. The simulation is replicated 500 times. For each iteration, the griddy-Gibbs sampler is employed with 5,000 draws (500 discarded) to generate the unrestricted elements of the cointegrating vector with the interval of integration (the deterministic Simpson's rule is used) for each element of $\beta_\star$ from -6.00 to 6.00 to avoid significant truncation of the posterior density.

Table 3.1 summarises the results of Monte Carlo simulations for testing for a Markov switching cointegration model against a linear cointegration/no cointegration model when transition probabilities are given by (p_{11}, p_{00})=(0, 1), (1, 0) and (0.90, 0.98). The values in the columns are the average posterior probabilities. When (p_{11}, p_{00})=(0, 1), the average posterior probability selects the correct model M0 with almost 100 percent even with T=100. When the true model is M1 with (p_{11}, p_{00})=(1, 0), the tests are quite powerful to detect the presence of cointegration while there is tendency to support M2. When the true model is M2 with (p_{11}, p_{00})=(0.90, 0.98), the tests tend to be somewhat conservative when the sample size is small with about 10 percent of the average posterior probability for M0 when T=100. With T=500, all probabilities for the correct model is almost 100 percent.

Next, we consider testing for Markov switching and model selection when the true model is the Markov switching cointegration model with two models specified M2 - M3 using the transition probabilities, $(p_{11}, p_{00}) \in \{(0.95, 0.90), (0.95, 0.95)\}$. The results are shown in Table 3.2. Power can be considered as the posterior probability for M1 versus the sum of the posterior probabilities of M2 and M3. Throughout the range of various models and sample sizes, the method has sufficient power to detect Markov switching behaviour. For model selection, the sample size T=100 is too small to detect the correct model when the true model is M3. Increasing the sample size to 200 improves the performances considerably.

Table 3.3 shows the Monte Carlo means and standard deviations for the estimated $\widehat{\beta}_\star = (\widehat{\beta}_2, \widehat{\beta}_3)$ for various models M1 - M3 when the true model is M2 and T=200 with (p_{11}, p_{00})=(0.95, 0.95) and the true values of β=(1, -1, 1) (thus, $\beta_\star$=(-1, 1)). If any Markov switching cointegration model (M2 and M3) is chosen, the cointegration vector can be accurately estimated. On the other hand, if the model is misspecified as a linear cointegration model (M1), the estimated $\widehat{\beta}$ can be greatly deviated from the true values and thus very unreliable with this sample size[2], hence the cointegrating vector should

[2]We run this experiments with the case when T = 500, and found the Monte Carlo

be estimated conditional on the regime variables.

means of β for M0 when the true model is M2 are much closer to the true values but still deviated with fairly large Monte Carlo standard deviations. With T=1000, those values are almost identical.

Table 3.1: Average posterior probabilities: Testing for Cointegration, Non-Cointegration and Markov Cointegration

True Model=M0: $(p_{11}, p_{00}) = (0, 1)$

model	$T = 100$	$T = 200$	$T = 500$
M0	0.996	0.998	0.999
M1	0.003	0.000	0.001
M2	0.001	0.002	0.000
M3	0.000	0.000	0.000

True Model=M1: $(p_{11}, p_{00}) = (1, 0)$

model	$T = 100$	$T = 200$	$T = 500$
M0	0.000	0.000	0.000
M1	0.824	0.912	0.999
M2	0.176	0.088	0.001
M3	0.000	0.000	0.000

True Model=M2: $(p_{11}, p_{00}) = (0.90, 0.98)$

model	$T = 100$	$T = 200$	$T = 500$
M0	0.107	0.042	0.000
M1	0.006	0.000	0.000
M2	0.788	0.913	1.000
M3	0.099	0.044	0.000

Table 3.2: Average posterior probabilities for model selection

$(p_{11}, p_{00}) = (0.95, 0.90)$ $(p_{11}, p_{00}) = (0.95, 0.95)$

True Model: M2

model	$T = 100$	$T = 200$	$T = 500$	$T = 100$	$T = 200$	$T = 500$
M0	0.000	0.000	0.000	0.000	0.000	0.000
M1	0.025	0.001	0.000	0.016	0.000	0.000
M2	0.896	0.968	0.994	0.943	0.980	0.971
M3	0.078	0.031	0.006	0.041	0.020	0.029

True model: M3

model	$T = 100$	$T = 200$	$T = 500$	$T = 100$	$T = 200$	$T = 500$
M0	0.017	0.000	0.000	0.020	0.000	0.000
M1	0.002	0.002	0.000	0.021	0.000	0.000
M2	0.346	0.103	0.001	0.316	0.101	0.002
M3	0.635	0.895	0.999	0.652	0.899	0.998

Table 3.3: Monte Carlo Means for $\widehat{\beta}_\star$ when the true model is M2 with T=200

(p_{11}, p_{00})=(0.95, 0.95)

model	$\overline{\widehat{\beta}_2}$	$\overline{\widehat{\beta}_3}$
M1	-0.536 (1.720)	-0.175 (1.884)
M2	-1.014 (0.226)	0.913 (0.364)
M3	-1.009 (0.110)	1.012 (0.312)

( ) is the Monte Carlo standard deviation

True values: $\beta_2 = -1$ and $\beta_3 = 1$

3.4 Application: PPP between UK and US

This section presents an application of the Markov switching cointegration model. We examine the simplest form of purchasing power parity between UK and US, using only three variables to see whether the cointegration would be switched off in one regime.

Whether PPP holds or not has been controversial among econometricians. Since Johansen and Juselius (1992) tested PPP and UIP for UK, many researchers have tried to find the cointegration relationship of PPP by adding other variables or using longer historical data[3]. Many theories for the flexible price monetary models, the behaviour of exchange rates, and the exchange rate target zone analysis assume that PPP always holds in the long-run. For example, Krugman (1988), Froot and Obstfeld (1989), Flood and Garber (1989), and Bertola and Caballero (1990). However, many empirical studies have shown that PPP has been rarely detected from Johansen and Juselius to a recent paper by Engel (2000), while some authors have found that the real exchange rates converge to their PPP level in the long-run, see Frankel (1986), Kim (1990), or Abuaf and Jorion (1990).

Even though PPP does not hold, cointegration among the three variable - exchange rate, CPIs of the two countries - is generally believed to exist. However, from the early 80's to the Plaza agreement (22th of September, 1985), the US dollar was highly over-valued with high interest rate in US. Then after the agreement the value of the dollar was corrected. These things might result in deviation from equilibrium. As Granger and Siklos (1996)

[3]See, for example, Froot, Kim and Rogoff (1995), who used 700 years of commodity price data.

pointed out, cointegration can be switched off because a common stochastic trend is added. In the simplest three-variable PPP context, the variable which could cause such a result is the interest rate.

In this section, we present an application of the Bayesian Markov switching cointegration model to a simple model of PPP for UK-US, where we expect from economic theory the cointegration relationship to be $e_t - p_t + p_t^{us}$, where e_t denotes the logarithm of exchange rate of the UK sterling against the US dollar, p_t the logarithm of UK CPI and p_t^{us} the logarithm of US CPI. Data used are quarterly with the range from the first quarter of 1961 to the first quarter of 1999 taken from the IMF's *International Financial Statistics.*

First, we consider testing for cointegrating rank for PPP relation of UK-US. The left side of Table 3.4 presents the results of cointegration tests by Bayesian posterior probabilities using the Bayes factor approximated by Schwarz BIC and the PIC of Phillips and Ploberger (1994, 1996) and a classical method by Johansen's LR trace statistics. The number of the lags in VAR is chosen as 2 based on the AIC and the Schwarz BIC. The results in Table 3.4 shows that each method selects different rank. The Bayes factor by the Schwarz BIC selects rank 0 with 100 percent of the posterior probability, while the PIC results in rank 2 with 99.1 percent. Johansen's trace test rejects the null of rank 0 with 5 percent significance level. If we assume that the rank is one, the estimated cointegrating vector β either by Bayesian or Johansen's method with a linear cointegration model is $\widehat{\beta}$=(1, -3.87, 5.53) and (1, -3.94, 5.61) respectively and these are far from the (1, -1, 1) that PPP implies. For Bayesian estimation of the cointegrating vector, β, we estimated using the method described in the previous chapter using the griddy-Gibbs sampler with 4,500 (plus 500 discarded) draws.

Combining the economic theory with the mixed test results, it seems reasonable to assume $r = 1$. Then, we proceed the model selection with model M0 - M3 described as in the previous section. We apply the Markov switching cointegration model with exactly the same prior specifications as in the previous section. The right side of Table 3.4 presents the posterior probabilities for each model for the UK-US PPPs. The Bayes factors (and thus posterior probabilities) show that there is strong support for nonlinearity against linear model M0 whose posterior probability is zero percent and that M2 (Markov model with a constant mean in VECM) is given dominantly highest posterior probability among other models and hence M2 is the most favoured among others. Thus we focus on M2 for investigating the posterior results.

Table 3.5 reports the posterior mean and standard deviation of each parameter and covariances for two regimes for M2 from Gibbs sampling output. Note that the standard deviations given by the Bayesian method are generally larger compared with the classical ML results, which are measured with the asymptotic distributions and thus give too optimistic results. See Bauwens and Lubrano (1999). From the results, the posterior mean of the cointegrating vector $\widehat{\beta}$ of M2 (= (1, -0.8329, 0.8745)) is much closer to the values that we expect from PPP and is significantly different from the estimated $\widehat{\beta}$ (= (1, -3.87, 5.53)) in the linear cointegration model. Thus, if we use $\widehat{\beta}$ based on the linear cointegration model, unconditional on the regime variables, as Balke and Fomby (1997) investigated the Markov switching cointegration model, results generated by the Markov switching cointegration models would be quite different.

Figure 3.1 presents the posterior expectation of the regime variable. Cointegration holds during early 60's to 72 of the fixed exchange rate regime. Introduction of float exchange rate and rapid inflation by two times oil shocks might have caused departure from the cointegration. In the first half of the 80s US dollar was misaligned and then the value of the dollar was corrected. The cointegration began to hold since 92' when the UK left the Exchange Rate Mechanism of the EMU.

Figure 3.2 shows the posterior densities of the cointegrating vector (left column) and the adjustment term (right column). The plots for the posterior densities of unrestricted elements of cointegrating vector $\beta_\star'$=(β_1, β_2) suggest that PPP relation, $\beta_\star'$= (-1, 1), is barely included in the densities region areas. The right column's plots for the posterior densities of the adjustment term show that zero does not lie within the density regions except for the first element α_1.

Table 3.4: Cointegration rank test (left) and Model Selection (right)

rank	Bayesian Pr (r | Y)	PIC Pr (r | Y)	Johansen trace test	null
0	1.000	0.000	47.91**	0
1	0.000	0.009	6.951	≤ 1
2	0.000	0.991	3.141	≤ 2
3	0.000	0.000	—	—

Model	Pr(M|Y)
M0	0.002
M1	0.000
M2	0.970
M3	0.028

Estimated cointegrating vector when $r = 1$
Bayesian: $\widehat{\beta}$=(1, -3.59, 5.53)
Johansen: $\widehat{\beta}$=(1, -3.94, 5.61)

Table 3.5: Posterior results for each parameter for M4

parameter	mean	s.d	parameter	mean	s.d.
β_2	-0.8329	0.0720	μ_1	0.0066	0.0046
β_3	0.8745	0.0899	μ_2	0.0220	0.0016
α_1	0.0317	0.0355	μ_3	0.0147	0.0008
α_2	0.0588	0.0227	p_{00}	0.9545	0.0249
α_3	0.0369	0.0141	p_{11}	0.9466	0.0288

$$\Sigma = \begin{bmatrix} 3.222e-03, & -6.135e-05, & -4.230e-05 \\ -6.135e-05, & 2.907e-04 & 8.526e-05 \\ -4.230e-05, & 8.526e-05 & 7.421e-05 \end{bmatrix},$$
$|\Sigma| = 4.574e-11,$

Figure 3.1: Posterior expectation of the regime variable E[S_t|Y] for UK/US PPP

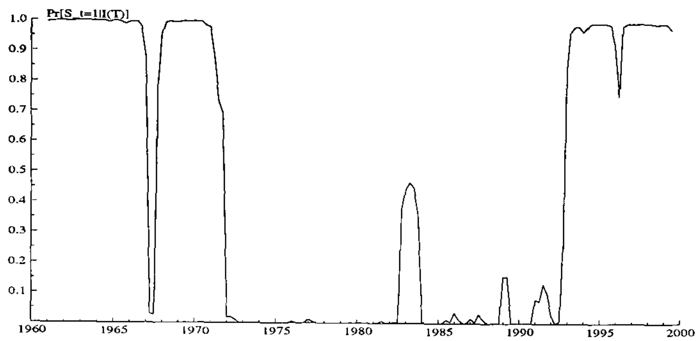
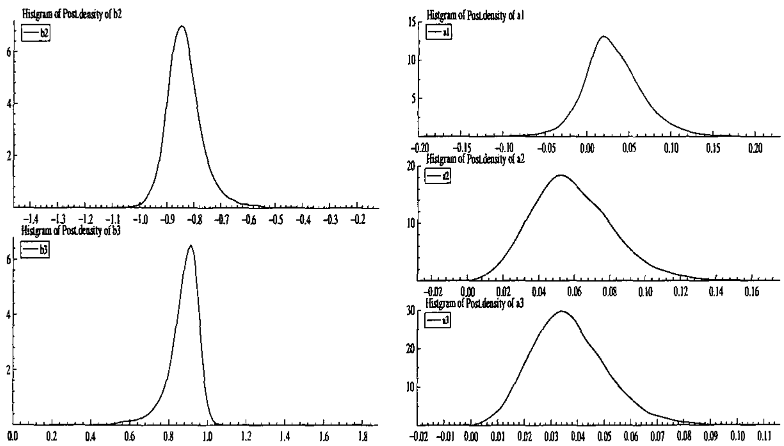


Figure 3.2: Histogram of Posterior Densities for $\beta_\star$ (left) and α (alpha) for UK/US PPP



3.5 Conclusion

In this chapter Bayesian inference for a Markov switching cointegration model is proposed. This model can be applied when researcher thinks the cointegration relationships contain a regime where disequilibrium occurs.

The estimations are carried out entirely by the Bayesian methods. The cointegrating vector is drawn in a nonlinear framework conditional on the regime variable in each iteration of the Gibbs sampling so that the estimation of the cointegrating vector is more accurate than the classical method by Balke and Fomby (1997) where the cointegrating vector is estimated unconditional on the regime variables, assuming the model is linear.

We have illustrated a Bayesian test for Markov switching behaviour and model selection. The Bayes factors provide comparisons with all the models under consideration, with either nested or non-nested models. Using Bayes factors can avoid the nonstandard inference problem in the classical methods. Thus, it is possible to compare a linear non-cointegrated model with nonlinear cointegrated models. Although we did not compare the proposed method with classical method by Hansen (1992 and 1996) to test for Markov switching behaviour, we believe that the proposed method has sufficient power to detect nonlinearity and the correct model specifications.

As an application for a Markov switching cointegration model, we examine PPP relation between UK and US. If we apply the Markov switching cointegration model, the cointegrating vector is much closer to the PPP, while linear cointegration model produces the cointegrating vector that is far from the PPP.

In this chapter Markov switching is chosen as a switching behaviour. It is, however, possible to consider alternative multivariate nonlinear models such as threshold or smooth transition models to analyse the nonlinear cointegration and compare these models with Markov switching model by the Bayes factors. Bayesian approach enables to modify easily to accommodate these models.

Chapter 4

Cointegrated Models with Structural Breaks

4.1 Introduction

In recent years, the econometrics literature on structural break in time series models has been extensively studied. Papers such as Perron (1989) have dealt with this issue in the framework of a priori imposed break dates, while others, such as Banerjee, Lumsdaine and Stock (1992), Christiano (1992) and Zivot and Andrews (1992), have used methods where the break date is endogenized. Much of the subsequent research has focused on testing for structural break when the break date may not be known. Among these, the *supF* statistic of Andrews (1993) and the *expF* and *aveF* statistics of Andrews-Ploberger (1994) are most notable. Hansen (2000) proposed a bootstrapping method for testing for a structural break, based on Andrews and Andrews-Ploberger's statistics.

An extension of the literature on testing for structural break involves allowing for more than one possible break date. For many macroeconomic or financial time series with the possibility of structural break, the assumption of at most one break date is unrealistic and restrictive. Bai and Perron (1998) proposed a test for multiple structural breaks at unknown dates using the double maximum test. Another testing method for detecting multiple changes includes Bai (1999), who proposed a likelihood ratio test with the null of l breaks against the alternative $l+1$ break points. While these methods only allow for a structural break in mean, breaks in variance are often found in economic and financial data. Schwert (1990) found that volatility of the stock-market is higher during and after the 1987 crash. Inclan (1993), Inclan and Tiao (1994), and Chen and Gupta (1997) detected multiple breaks in variance for various series of stock returns. Engel and Hakkio (1996) found that European Monetary System exchange rates have higher volatility during the periods of alignment, and Kim and Engel (1999) found multiple breaks in variance in real exchange rates associated with historically significant monetary events. Kim and Nelson (1999) combined structural break with Markov switching model to find evidence of variance breaks in postwar business cycles.

The above literature considered structural break(s) in univariate models. Recently, some researchers have considered estimation of and testing for structural break in multivariate cointegrated models. Gregory and Hansen (1996a) studied residual-based tests for cointegration with a single structural break. They proposed $ADF-$, $Z_\alpha-$, and Z_t-type tests designed to test the null of no cointegration against the alternative of cointegration in the presence of a possible regime shift. Gregory and Hansen (1996b) extended this work, by permitting a trend shift as well as a regime shift and provide the critical values for testing cointegration with one break. Seo (1998) derived the Lagrange multiplier test for structural breaks in cointegration relations and adjustment terms, using the framework of Andrews and Ploberger (1994). Hansen and Johansen (1999) also tested parameter instability in cointegrating vectors based on Nyblom's L statistic (1989). Hansen (2003) explored the multiple-break case in cointegrated systems, and allowed changes in any subset of the parameters, when the time of the change points and the number of cointegration relations are treated as known. Inoue (1999) derived a rank test for cointegrated systems with a structural change in trend.

In this chapter, we investigate multiple structural breaks in the level, trend and covariance of a co-integrated VAR model, using Bayesian approach which extends Wang and Zivot (2000)'s method for detecting multiple structural changes in univariate models. We consider structural breaks in the level, trend and error covariance, but also present a case that structural breaks occur in the adjustment term and the cointegrating vector. Just a slight modification will make it possible to extend our method to detect breaks in any subset of the parameters in the cointegrated system. Hansen (2003) considered the similar general cointegration models with structural breaks in any subset of parameters, however the location of the break points are assumed known. There is no research paper that deals with general cointegration model with structural breaks in any subset of the parameters where the break points are unknown.

We assume that cointegration rank does not change with structural breaks, though it is not difficult to test for the cointegration rank in the presence of breaks simply by using methods proposed by Kleibergen and Paap (2002), Chao and Phillips (1999)'s PIC, or the method that introduced in Chapter 2, by conditioning on the structural breaks, while classical methods require substantial modification, since the case with unknown change points leads to non-standard asymptotic distributions.

The Bayesian approach has several advantages over the classical method as in the context of structural break models as it is technically simpler, allows inferences that are optimal given the framework, and allows for nonnested model comparison by computing posterior odds (see Raftery (1994)). Additionally, inference from Bayesian approach is based on the exact finite sample properties for all of the parameters of the model including the break dates. Finally, unlike most classical methods for detecting structural breaks, the Bayesian approach provides information of uncertainty in the location of the break dates. If the posterior probability mass function of the change point exhibits a substantial range in dates, we might consider that the structural break occurs smoothly, rather than suddenly at one particular date. In order to determine the number of structural breaks, we take it as a model selection problem from the posterior odds by Bayes factors.

This chapter is organised as follows. Section 4.2 presents the statistical co-integrated model with multiple structural breaks in the level, trend and error covariance. We also present a case of the breaks in the adjustment term and the cointegrating vector. Then, we specify the prior densities for all parameters, likelihood function, and the posterior densities. Section 4.3 provides the issue of testing or model selection for detecting multiple structural breaks using Bayes factors. Monte Carlo simulations using artificially generated data are presented in Section 4.4. We apply the method to investigate Japanese term structure of interest rates in Section 4.5. Section 4.6 concludes. All computation in this chapter are performed using code written by the author with Ox v3.30 for Linux (Doornik, 1998).

4.2 A Time Series Model with Multiple Structural Breaks in Co-integrated VAR Model

4.2.1 Statistical Model

To investigate a co-integrated multivariate model with multiple structural breaks, we consider the form of a vector error correction model with multiple structural breaks in level, trend and error covariance. Let X_t denote an $I(1)$ vector of n-dimensional time series with r linear cointegrating relations. The long-run multiplier matrix is decomposed as $\alpha\beta'$, both are $n \times r$, where α is the adjustment term and β' is the cointegrating vector. If we assume that the intercept term μ, trend ξ and error covariance σ_t in VECM are subject to structural breaks, then the VECM representation is:

$$\Delta X_t = \mu_t + \xi_t t + \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \sigma_t \varepsilon_t \tag{4.1}$$

where $t = p, p+1, \ldots, T$, and p is the number of lags, and ε_t are assumed $N(0, I_n)$ and independent over time. Dimensions of matrices are μ_t, ξ_t and ε $(n \times 1)$, Ψ and σ_t $(n \times n)$. We assume that the parameters μ_t, ξ_t and σ_t

are subject to $m < t$ structural breaks with break points $k_1, \ldots, k_m$, where $k_1 < k_2 < \cdots < k_m$, so that the observations can be separated into $m+1$ regimes.

Equation (4.1) can be rewritten in the matrix format as:

$$Y = X\Gamma + Z\beta\alpha' + E = WB + E \tag{4.2}$$

where

$$Y = \begin{bmatrix} \Delta X_p' \\ \Delta X_{p+1}' \\ \vdots \\ \Delta X_T' \end{bmatrix}, Z = \begin{bmatrix} X_{p-1}' \\ X_p' \\ \vdots \\ X_{T-1}' \end{bmatrix}, E = \begin{bmatrix} \varepsilon_p'\sigma_p' \\ \varepsilon_{p+1}'\sigma_{p+1}' \\ \vdots \\ \varepsilon_T'\sigma_T' \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} \mu_1 & \cdots & \mu_{m+1} & \xi_1 & \cdots & \xi_{m+1} & \Psi_1 & \cdots & \Psi_{p-1} \end{bmatrix},$$

$$X = \begin{bmatrix} s_{1,p} & \cdots & s_{m+1,p} & s_{1,p} & \cdots & s_{m+1,p} & \Delta X_{p-1}' & \cdots & \Delta X_1' \\ s_{1,p+1} & \cdots & s_{m+1,p+1} & 2s_{1,p+1} & \cdots & 2s_{m+1,p+1} & \Delta X_p' & \cdots & \Delta X_2' \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{1,T} & \cdots & s_{m+1,T} & (T-p+1)s_{1,T} & \cdots & (T-p+1)s_{m+1,T} & \Delta X_{T-1}' & \cdots & \Delta X_{T-p+1}' \end{bmatrix}$$

$$W = \begin{bmatrix} X & Z\beta \end{bmatrix}, B = \begin{bmatrix} \Gamma' & \alpha \end{bmatrix}'$$

Let τ be the number of rows of Y, so that $\tau = T - p + 1$, then X is $\tau \times (2(m+1) + n(p-1))$, Γ is $(2(m+1) + n(p-1)) \times n$, W is $\tau \times k$ where $k = 2(m+1) + n(p-1) + r$, and B is $k \times n$. $s_{i,j}$ in X is an indicator variable which equals to 1 if regime is i and 0 otherwise. Equation (4.2) represents the multivariate regression format of (4.1).

If we wish to consider a cointegration model with multiple structural breaks not only in the level, trend and error covariance but also in the adjustment term α and/or the cointegrating vector β as:

$$\Delta X_t = \mu_t + \xi_t t + \alpha_t \beta_t' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \sigma_t \varepsilon_t, \tag{4.3}$$

we simply modify W, Z and B in (4.2) as:

$$W = \begin{bmatrix} X & Z_1\beta_1 & \cdots & Z_{m+1}\beta_{m+1} \end{bmatrix}, \; B = \begin{bmatrix} \Gamma' & \alpha_1 & \cdots & \alpha_{m+1} \end{bmatrix}',$$

$$Z_i = \begin{bmatrix} s_{i,p}X_{p-1}' \\ s_{i,p+1}X_p' \\ \vdots \\ s_{i,T}X_{T-1}' \end{bmatrix} \text{ for } i = 1, \ldots, m+1.$$

so that we can easily consider a more general cointegration model with structural breaks in the level, trend, error covariance, adjustment term and the cointegrating vectors. It is also possible to consider that the lag terms are subject to the structural breaks. However, for simplicity, we focus on the cointegration model with breaks in the level, trend and covariance in this thesis.

4.2.2 Prior Distributions and Likelihood Functions

Let $k = (k_1, k_2, \ldots, k_m)'$ denote the vector of break dates. For the prior for the location of the break dates k, we choose a diffuse prior such that the prior is discrete uniform over all ordered subsequences of $t = 2, 3, \ldots, T$. We specify a Student-t prior for β, and the natural conjugate priors for other parameters are chosen, assuming prior independence between k, B and Σ_i, $i = 1, 2, \ldots, m+1$, such that $p(k, B, \beta_\star, \Sigma_1, \Sigma_2, \ldots, \Sigma_{m+1}) = p(k)\, p(B)\, p(\beta_\star) \prod_{i=1}^{m+1} p(\Sigma_i)$. This is because if we consider the prior density for B is conditional on Σ as is often used in regression models as in the two previous chapters, it is not convenient to consider a case when the error covariance is also subject to structural breaks. Thus, we consider that the prior density for B is the marginal distribution unconditional on Σ and vectorize B as follows:

$$p(k) \propto 1_{\{t=2:t=T\}} \tag{4.4}$$

$$\Sigma_i \sim IW(J_i, h) \tag{4.5}$$

$$vec(B) \sim N(vec(B_0), \Sigma_B) \tag{4.6}$$

$$\beta_\star \sim Mt\left(\overline{\beta_\star}, H, Q, v\right) \tag{4.7}$$

where IW refers to an inverted Wishart distribution with parameters $J_i \in \mathbb{R}^{n\times n}$ and degrees of freedom, h; N refers to a multivariate normal with mean $vec(B_0) \in \mathbb{R}^{kn\times 1}$, $k = m + n(p-1) + mr$ and covariance $\Sigma_B \in \mathbb{R}^{kn\times kn}$; Mt refers to a matrix Student-t distribution with parameters $\overline{\beta} \in \mathbb{R}^{(n-r)\times r}$, $Q \in \mathbb{R}^{n-r}$, $H \in \mathbb{R}^{r}$. Note that r^2 restrictions for identification are imposed on β such that $\beta' = \begin{pmatrix} I_r & \beta_\star' \end{pmatrix}$, where $\beta_\star$ is $(n-r)\times r$ unrestricted matrix as shown in the previous chapters.

The joint prior of k, B, Σ_i and $\beta_\star$ is given by multiplication of (4.4) - (4.7) as follows:

$$
\begin{aligned}
&p\left(k, B, \beta_{\star}, \Sigma_1, \Sigma_2, \ldots, \Sigma_{m+1}\right) \\
&\propto g\left(\beta_{\star}\right)\left|\Sigma_B\right|^{-1/2}\left(\prod_{i=1}^{m+1}\left|J_i\right|^{h/2}\left|\Sigma_i\right|^{-1/2}\right) \\
&\times \exp\left[-\frac{1}{2}\left\{\operatorname{tr}\left[\sum_{i=1}^{m+1}\left(\Sigma_i^{-1} J_i\right)\right]+vec\left(B-B_0\right)' \Sigma_B^{-1} vec\left(B-B_0\right)\right\}\right]
\end{aligned} \tag{4.8}
$$

where $g\left(\beta_{\star}\right)$

The likelihood function for $k, B, \Sigma_1, \ldots \Sigma_{m+1}$, β is given by,

$$
\begin{aligned}
&\mathfrak{L}\left(k, B, \beta, \Sigma_1, \ldots, \Sigma_{m+1} \mid Y\right) \\
&\propto\left(\prod_{i=1}^{m+1}\left|\Sigma_i\right|^{-t_i/2}\right) \exp\left(-\frac{1}{2} \operatorname{tr}\left[\sum_{i=1}^{m+1}\left\{\Sigma_i^{-1}\left(Y_i-W_i B\right)'\left(Y_i-W_i B\right)\right\}\right]\right) \\
&=\left(\prod_{i=1}^{m+1}\left|\Sigma_i\right|^{-t_i/2}\right) \exp\left(-\frac{1}{2} \sum_{i=1}^{m+1}\left[\left(vec\left(Y_i-W_i B\right)\right)'\left(\Sigma_i \otimes I_T\right)^{-1}\left(vec\left(Y_i-W_i B\right)\right)\right]\right)
\end{aligned} \tag{4.9}
$$

where Y_i denotes the $t_i \times n$ matrix of Y values in regime i, W_i denotes $t_i \times 2(m+1)+n(p-1)+mr$ matrix of W values in regime i, and t_i is the number of observations in regime i when $s_t = i$, $i = 1, 2, \ldots, m+1$.

4.2.3 Posterior Specifications and Estimation

Now, the joint posterior distribution can be obtained from the joint priors given in (4.8) and the likelihood function for k, B, Σ, and β that is,

$$
\begin{aligned}
&p\left(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star \mid Y\right) \propto p\left(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star\right) \mathfrak{L}\left(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star \mid Y\right) \\
&\propto g\left(\beta_\star\right) \left|\Sigma_B\right|^{-1/2} \left(\prod_{i=1}^{m+1} \left\{\left|J_i\right|^{h/2} \left|\Sigma_i\right|^{-(t_i-1)/2}\right\}\right) \\
&\times \exp\left(-\frac{1}{2}\left[\operatorname{tr}\left(\sum_{i=1}^{m+1} \Sigma_i^{-1}\right) + \sum_{i=1}^{m+1} \left\{\left(\left[vec\left(Y_i - W_i B\right)\right]' \left(\Sigma_i \otimes I_\tau\right)^{-1} vec\left(Y_i - W_i B\right)\right)\right\}\right.\right. \\
&\left.\left. + vec\left(B - B_0\right)' \Sigma_B^{-1} vec\left(B - B_0\right)\right]\right) \qquad (4.10)
\end{aligned}
$$

where $g\left(\beta_\star\right)$ refers to the prior for β given in (4.7). Consider first the conditional posterior of k_i, $i = 1, 2, \ldots, m$. Given that $1 = k_0 < \cdots < k_{i-1} < k_i < k_{i+1} < \cdots < k_{m+1} = \tau$ and the form of the joint prior, the sample space of the conditional posterior of k_i only depends on the neighbouring break dates k_{i-1} and k_{i+1}. It follows that, for $k_i \in [k_{i-1}, k_{i+1}]$,

$$
p\left(k_i \mid B, \Sigma_i, \beta_\star, Y_i\right) \propto p\left(k_i \mid k_{i-1}, k_{i+1}, B, \beta, \Sigma_i, Y_i\right) \qquad (4.11)
$$

for $i = 1, \ldots m$, which is proportional to the likelihood function for $\Theta = (B, \beta, \Sigma_i)$ evaluated with a break at k_i only using data between k_{i-1} and k_{i+1} and probabilities proportional to the likelihood function.

Next, we consider the conditional posterior of Σ_i, $vec\left(B\right)$ and $\beta_\star$. From (4.10), we can write two terms as:

$$
\begin{aligned}
&\sum_{i=1}^{m+1} \left\{\left[vec\left(Y_i - W_i B\right)\right]' \left(\Sigma_i \otimes I_\tau\right)^{-1} vec\left(Y_i - W_i B\right)\right\} \\
&+ \left[vec\left(B - B_0\right)\right]' \Sigma_B^{-1} vec\left(B - B_0\right)
\end{aligned}
$$

$$= [vec\,(B - B_\star)]'\, M_\star^{-1} vec\,(B - B_\star) + Q$$

where

$$Q = \sum_{i=1}^{m+1} \left\{ [vec\,(Y_i)]' \left(\Sigma_i \otimes I_T\right)^{-1} vec\,(Y_i) \right\} + [vec\,(B_0)]'\, \Sigma_B^{-1} vec\,(B_0) - [vec\,(B_\star)]'\, M_\star^{-1} vec\,(B_\star)$$

Thus, the conditional posterior of Σ_i is derived as an inverted Wishart distribution as:

$$p(\Sigma_i \mid k, B, \beta_\star, Y) \propto |S_{i,\star}|^{t_i/2} |\Sigma_i|^{-(t_i+h+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_i^{-1} S_{i,\star}\right)\right] \qquad (4.12)$$

where $S_{i,\star} = (Y_i - W_i B)'\,(Y_i - W_i B) + J_i$. The conditional posterior of $vec\,(B)$ as a multivariate normal density with covariance, $M_\star$, that is,

$$p(vec\,(B) \mid k, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star, Y) \propto |\Sigma_B|^{-1/2} \exp\left[-\frac{1}{2}\left\{[vec\,(B - B_\star)]'\, M_\star^{-1} vec\,(B - B_\star)\right\}\right] \qquad (4.13)$$

where

$$vec\,(B_\star) = \left[\Sigma_B^{-1} + \sum_{i=1}^{m+1}\left\{\Sigma_i^{-1} \otimes (W_i' W_i)\right\}\right]^{-1} \left[\Sigma_B^{-1} vec\,(B_0) + \sum_{i=1}^{m+1}\left\{\left(\Sigma_i \otimes I_k\right)^{-1} vec\,(W_i' Y_i)\right\}\right],$$

and

$$M_\star = \left[\Sigma_B^{-1} + \sum_{i=1}^{m+1}\left\{\Sigma_i^{-1} \otimes (W_i' W_i)\right\}\right]^{-1}$$

The posterior of $\beta_\star$ can be derived by integrating (4.10) with respect to Σ_i and B, but we choose the other way to derive it. The prior of B in (4.6) can be written as a matric normal as:

$$p(B) \propto \mid \Sigma_{B2} \mid^{-k/2} \mid \Phi \mid^{n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma_{B2}^{-1}(B-B_0)'\,\Phi^{-1}(B-B_0)\right\}\right], \tag{4.14}$$

where $\Sigma_{B2} \otimes \Phi = \Sigma_B$, Σ_B is $n \times n$, Φ is $k \times k$. Then the likelihood in (4.9) can be written as:

$$\begin{aligned}\mathfrak{L} &\propto \left(\prod_{i=1}^{m+1} |\Sigma_i|^{-t_i/2}\right) \exp\left[-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{m+1}\left\{\Sigma_i^{-1}(Y_i - W_iB)'(Y_i - W_iB)\right\}\right)\right] \\ &\propto \left(\prod_{i=1}^{m+1} |\Sigma_i|^{-t_i/2}\right) \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{m+1}\left[\Sigma_i^{-1}\left\{\widehat{S}_i + \left(B-\widehat{B}_i\right)' W_i'W_i\left(B-\widehat{B}_i\right)\right\}\right]\right)\right\}\end{aligned} \tag{4.15}$$

where $\widehat{B}_i = (W_i'W_i)^{-1}W_i'Y_i$ and $\widehat{S}_i = \left(Y_i - W_i\widehat{B}_i\right)'\left(Y_i - W_i\widehat{B}_i\right)$. Multiplying the likelihood in (4.15) by the prior densities (4.14), (4.4), (4.5), and (4.7), we have the joint posteriors as:

$$\begin{aligned}&p(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star \mid Y) \propto p(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star)\,\mathfrak{L}(k, B, \Sigma_1, \ldots, \Sigma_{m+1}, \beta_\star \mid Y) \\ &\propto g(\beta_\star)\prod_{i=1}^{m+1}\left(|J_i|^{h/2}\left|\Sigma_i^{-(h+n+1)/2}\right|\right)|\Sigma_{B2}|^{-k/2} \\ &\times\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{m+1}\left[\Sigma_i^{-1}\left\{J_i + \widehat{S}_i + \left(B-\widehat{B}_i\right)' W_i'W_i\left(B-\widehat{B}_i\right)\right\}\right]\right.\right.\end{aligned}$$

$$+\Sigma_{B2}^{-1}\left(B-B_0\right)'\Phi\left(B-B_0\right))\} \tag{4.16}$$

Integrating the joint posterior in (4.16) with respect to $\Sigma_1, \ldots, \Sigma_{m+1}$, we have the joint posterior of B and $\beta_\star$ as follows:

$$\begin{aligned} p\left(B,\beta_\star|k,Y\right) &\propto \int\int\cdots\int p\left(B,\Sigma_1,\Sigma_2,\ldots,\Sigma_{m+1},\beta_\star\right)d\Sigma_1 d\Sigma_2\ldots d\Sigma_{m+1} \\ &\propto g\left(\beta_\star\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left[\sum_{i=1}^{m+1}\left\{J_i+\widehat{S}_i+\left(B-\widehat{B}_i\right)'W_i'W_i\left(B-\widehat{B}_i\right)\right\}\right]\right), \end{aligned} \tag{4.17}$$

By integrating (4.17) with respect to B we obtain the posterior density of the cointegrating vector $\beta_\star$ as follows:

$$\begin{aligned} p\left(\beta_\star|k,Y\right) &\propto \int p\left(B,\beta_\star \mid k,Y\right)\mathrm{d}B \\ &\propto \int g\left(\beta_\star\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left[\sum_{i=1}^{m+1}\left\{J_i+\widehat{S}_i+\left(B-\widehat{B}_i\right)'W_i'W_i\left(B-\widehat{B}_i\right)\right\}\right]\right)\mathrm{d}B \\ &\propto g\left(\beta_\star\right)\prod_{i=1}^{m+1}\left(\left|J_i+\widehat{S}_i\right|^{-(t-k)/2}\left|W_i'W_i\right|^{-n/2}\right), \end{aligned} \tag{4.18}$$

If diffuse priors for Σ_i are used ($J_i = 0$ and $h = 0$) instead of the inverted Wishart, the posterior of $\beta_\star$ can be simplified as:

$$p\left(\beta_\star|k,Y\right) \propto g\left(\beta_\star\right)\prod_{i=1}^{m+1}\frac{|\beta_\star' W_i^0\beta_\star|^{l_i^0}}{|\beta_\star' W_i^1\beta_\star|^{l_i^1}},$$

where

$$W_i^0 = Z_i' M_i^X Z_i, \quad M_i^X = I_{t_i} - X_i' \left(X_i' X_i\right)^{-1} X_i', \quad M_i^Y = I_{t_i} - Y_i' \left(Y_i' Y_i\right)^{-1} Y_i,$$
$$W_i^1 = Z_i' M_i^Y \left[I_{t_i} - X_i \left(X_i' M_i^Y X_i\right)^{-1} X_i'\right] M_i^Y Z_i = Z_i' M_i^X \left[I_{t_i} - Y_i \left(Y_i' M_i^X Y_i\right)^{-1} Y_i'\right] M_i^X Z_i,$$
$$l_i^0 = (t_i - k - n)/2, \quad l_i^1 = (t_i - k)/2.$$

This is a result of slight modification of the Theorem 9.3 in Bauwens, *et al* (1999), which shows that the posterior of β belongs to a matric-variate generalization of the class of poly-t densities (poly-matrix-t).

To summarise, we have following conditional posterior densities as:

$$\Sigma_i \mid k, \beta, B, Y \sim IW\left(\left(Y_i - W_i B\right)' \left(Y_i - W_i B\right) + J_i, t_i + h\right) \text{ for } \forall i \tag{4.19}$$

$$vec\left(B\right) \mid k, \Sigma_1, \ldots, \Sigma_{m+1}, \beta, Y \sim N\left(vec\left(B_\star\right), M_\star\right) \tag{4.20}$$

$$p(\beta_\star \mid k, Y) \propto g\left(\beta_\star\right) \prod_{i=1}^{m+1} \left(\left|J_i + \widehat{S}_i\right|^{-(t-k)/2} \left|W_i' W_i\right|^{-n/2}\right) \tag{4.21}$$

The posterior distributions obtained in (4.19), (4.20) and (4.21) are not convenient analytical forms. Rather they are conditional on other parameters which must be estimated. A Gibbs sampler can be employed to generate random draws from the conditional posteriors. While the conditional posterior densities for Σ_i and B are of a known form, the posterior for $\beta_\star$ in (4.21) is not a known form and thus can be drawn by employing Importance sampling, the Metropolis-Hastings algorithm (see Chib and Greenberg, 1995) or the Griddy-Gibbs sampling (see Ritter and Tanner, 1992). In this chapter, as chosen in the previous chapters, we chose the Griddy-Gibbs sampling technique because the algorithm does not require the specification of some function that approximates the distribution. Choosing the Griddy-Gibbs sampler, however, requires the appropriate choice of the grid of points and the computing cost is much higher than for the other two algorithms.

Given the full set of conditional posterior specifications above, we illustrate the Gibbs sampling algorithm for generating sample draws from the joint posterior. The following steps can be replicated:

- Step 1: Set $j = 1$. Specify starting values for the parameters of the model, $k^{(0)}$ $B^{(0)}$, $\beta^{(0)}$ and $\Sigma_i^{(0)}$, where Σ_i is a covariance matrix at regime i.

- Step 2a: Compute likelihood probabilities sequentially for each date at $k_1 = k_0^{(j-1)} + 1, \ldots, k_2^{(j-1)} - 1$ to construct a multinomial distribution. Weight these probabilities such that the sum of them equals 1.

- Step 2b: Generate a draw for the first break date k_i as a multinomial random variable on the sample space $\left[k_0^{(j-1)}, k_2^{(j-1)}\right]$ from

$$p\left(k_1^{(j)} \mid k_0^{(j-1)}, k_2^{(j-1)}, B^{(j-1)}, \beta^{(j-1)}, \Sigma_i^{(j-1)}, Y\right)$$

  and set $k_i = k_1$.

- Step 2c ($i = 3, \ldots, m+1$): Generate a draw of the $(i-1)$th break date $k_{i-1}^{(j)}$ from the conditional posterior $p\left(k_{i-1}^{(j)} \mid k_{i-2}^{(j-1)}, k_i^{(j-1)}, B^{(j-1)}, \beta^{(j-1)}, \Sigma_i^{(j-1)}, Y\right)$. Go back to Step 2a, but with imposing previously generated break date, in order to generate next break date. Iterate until all breaks are generated.

- Step 3: Generate $\beta_\star^{(j)}$, the unrestricted elements of β, from $p(\beta_\star \mid k, Y)$ using the Griddy Gibbs sampling algorithm (or Metropolis-Hastings/Importance sampling algorithm).

- Step 4: Generate $B^{(j)}$ from $p(vec\,(B) \mid k^{(j)}, \beta_\star^{(j)}, \Sigma_i^{(j-1)}, \ldots \Sigma_{m+1}^{(j-1)}, Y)$.

- Step 5: Generate $\Sigma_i^{(j)}$ from $p(\Sigma_i \mid k^{(j)}, \beta_\star^{(j)}, B^{(j)}, Y)$ for all $i = 1, \ldots, m+1$.

- Step 6: Set $j = j+1$, and go to Step 2.

Step 2 through to Step 6 can be iterated N times to obtain the posterior densities. Note that the first L iterations are discarded in order to remove the effect of the initial values.

4.3 Testing for Structural Break and Model Selection by Bayes Factors

In this chapter we deal with testing for multiple structural breaks in a vector error correction model as a problem of model selection and approximate the Bayes factors by Schwarz's Bayes information criterion (BIC) to select the the number of structural breaks. The Schwarz BIC can give a rough approximation to the Bayes factors, which is easy to use and does not require evaluation of the prior distribution, as Kass and Raftery (1995) noted. Wang and Zivot (2000) employ the Schwarz BIC to compute the Bayes factors for detecting the number of structural breaks in a univariate context. The Schwarz BIC for model j can be computed as

$$\mathrm{BIC}_j = -2\ln\mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right) + q_j \ln(t) \tag{4.22}$$

where $\mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right)$ denotes the likelihood function for model j; q_j denotes the total number of estimated parameters in the model j and M_j denotes the model indicator for model j. The likelihood function $\mathfrak{L}\left(\widehat{\theta_j} \mid Y; \mathrm{M}_j\right)$ is evaluated at $\widehat{\theta_j}$, the posterior means of the parameters for model j.

The Bayes factor for model k against model j can be approximated by

$$BF_{jk} = \exp\left[0.5\left(\mathrm{BIC}_j - \mathrm{BIC}_k\right)\right] \tag{4.23}$$

With the prior odds, defined as $\Pr(\mathrm{M}_k)/\Pr(\mathrm{M}_j)$, the posterior odds can be computed by multiplying the Bayes factor by the prior odds as $\mathrm{PosteriorOdds}_{jk} = BF_{jk} \times \mathrm{PriorOdds}_{jk}$. We compute the posterior odds for all possible models and then obtain the posterior probability for each model by computing

$$\Pr\left(\mathrm{M}_j \mid Y\right) = \frac{\mathrm{PosteriorOdds}_{jk}}{\sum_{m=1}^{n} \mathrm{PosteriorOdds}_{mj}} \tag{4.24}$$

where n is the number of models we consider.

By using the Schwarz BIC to approximate to logarithm of the Bayes factor, it is easy to test for the number of breaks and other model specification such as whether the error covariance is subject to structural breaks, as a problem of model selection. In our case, we compute the Schwarz BIC as

$$\mathrm{BIC}_j = -2\ln \mathfrak{L}\left(k_j, B, \Sigma_1, \Sigma_2, \ldots, \Sigma_{m+1}, \beta_\star \mid Y; \mathrm{M}_j\right) + q_j \ln(t) \quad (4.25)$$

We compute BIC(m) using the posterior modes of k_i for $i = 1, \ldots, m$ and the posterior means of the remaining parameters based on the output of the Gibbs sampler.

4.4 Simulation

In this section, we conduct a small Monte Carlo simulation of the approach outlined in Section 4.2 in order to examine its performance. We consider three data generation processes (DGPs) of a two-variable co-integrated model:

$$\begin{aligned}
&\text{DGP1} \quad \Delta y_t = \mu + \gamma t + \alpha\beta' y_{t-1} + \sigma\varepsilon \\
&\text{DGP2} \quad \Delta y_t = \mu_t + \gamma_t t + \alpha\beta' y_{t-1} + \sigma\varepsilon \qquad (4.26)\\
&\text{DGP3} \quad \Delta y_t = \mu_t + \gamma_t t + \alpha\beta' y_{t-1} + \sigma_t\varepsilon_t
\end{aligned}$$

DGP1 represents a no structural break model, DGP2 represents a structural break model in the level and trend, and DGP3 represents a structural break model in the mean, trend, and error covariance. The sample size in this experiment is 300 for all DGPs. We assume that there are two structural breaks for DGP2 and DGP3 at $t = 100$ and 200. The parameters in (4.26) are given as follows: $\mu = (-0.1261, 0.0677)'$, $\gamma = (0.0001, 0.0001)'$, $\alpha = (-0.0535, 0.0575)'$, $\beta = (1, -1.2764)'$, $\Sigma = \begin{bmatrix} 0.0471 & 0.0144 \\ 0.0144 & 0.0712 \end{bmatrix}$ for DGP1. The parameters for DGP2 and DGP3 are given in Table 4.2 and 4.3

respectively. These DGPs are intended to mimic the behaviour of Japanese interest rates. Figure 4.1 shows a simulated series from DGP2. It shows that the two-series exhibit two possible structural breaks, though it is ambiguous exactly when the breaks occurred.

The Gibbs sampling algorithm presented in Section 4.2 is implemented for the estimation of models for $m = 0, 1, \ldots, 5$ break points. For prior parameters, we set $J_i = I_3/1000$ for covariance prior in (4.5), $\Sigma_B = I_k/1000$ in (4.6) and $H = 1/1000$ in (4.7) to ensure fairly large variance for representing prior ignorance, $\overline{\beta}_\star = -1$, $v = 2.001$, $Q = I_3$, $B_0 = 0$. The number of cointegration rank and the number of the lags in VAR are assumed to be known. Also, we assume that correct model specifications are known for each model except the number of breaks. We assign an equal prior probability to each model with i breaks, so that $\frac{Pr(M_i)}{Pr(M_0)} = 1$[1]. After running the Gibbs sampler for 500 iterations, we save the next 2,000 draws for inference. This procedure is replicated 100 times. For each iteration , the Griddy-Gibbs sampler is employed to generate the unrestricted elements of the cointegrating vector with the interval of integration (the deterministic Simpson's rule is used) for each elements of $\beta_\star$ from -5.00 to 5.00 with 1,000 grid points.

Table 4.1 summarises the results of the Monte Carlo simulations for model selection. Each element in the Table shows the average posterior probability out of 500 replications. We compute the posterior probability using (4.23), (4.24) and (4.25). For the no break case (DGP 1), the correct model with $m = 0$ is chosen 91.4% of the time. For cointegrated models with breaks

[1]Inclan (1993) and Wang and Zivot (2000) used the prior odds as an independent Bernoulli process with probability $p \in [0, 1]$.

Figure 4.1: An Example of a Simulated Series from DGP2

(DGP 2 and 3), there is always very strong evidence in favour of $m = 2$. As we expected, DGP 3 shows the best performance among others with 96.8% of the time for $m = 2$.

Table 4.2 and 4.3 report the Monte Carlo means of the posterior means and the Monte Carlo standard deviations of the posterior means for DGP1 and DGP2 respectively. These tables show the estimation performance when the number of structural breaks is correctly specified.

The Gibbs sampler also provides the posterior mass function for each estimated break point. The posterior mass function of the break points for DGP 2 are plotted in Figure 4.2 for the series generated in the Monte Carlo study. The posterior mass function of each break points has a mode at the true break date.

Figure 4.2: Posterior Probability Mass of the Break Dates - An Example from DGP2

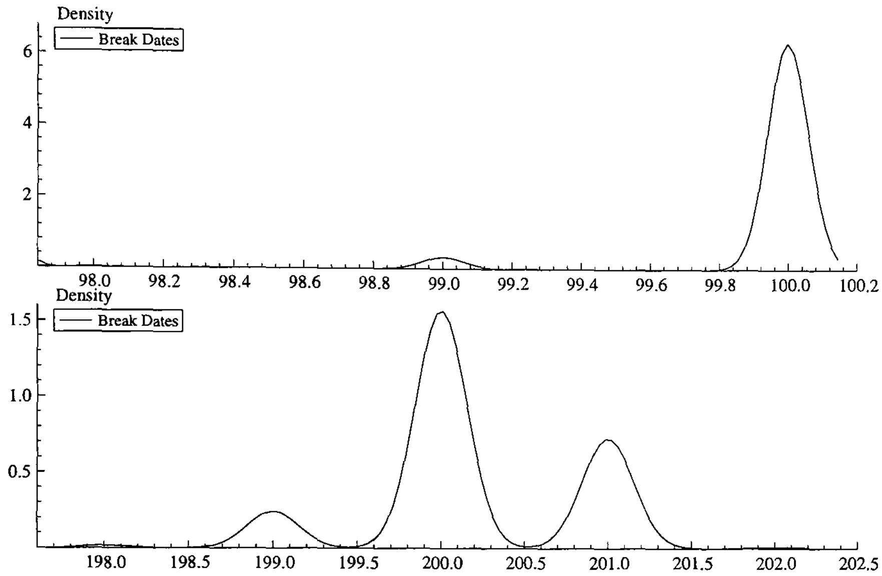


Table 4.1: Monte Carlo Results: Average posterior probabilities

DGP	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
DGP 1	0.914	0.085	0.001	0.000	0.000	0.000
DGP 2	0.003	0.026	0.847	0.123	0.001	0.000
DGP 3	0.000	0.007	0.968	0.045	0.000	0.000

Table 4.2: DGP2 - Estimation Results When $m = 2$

Parameter	True values	MC mean	MC SD
μ_{1-1}	-0.0606	-0.0549	0.0255
μ_{1-2}	-0.0909	-0.0869	0.0301
μ_{2-1}	0.0137	0.0220	0.0063
μ_{2-2}	-0.0481	-0.0393	0.0239
μ_{3-1}	0.0015	0.0023	0.0012
μ_{3-2}	-0.0086	-0.0175	0.0101
γ_{1-1}	0.0028	0.0027	0.0007
γ_{1-2}	0.0016	0.0013	0.0006
γ_{2-1}	-0.0013	-0.0010	0.0011
γ_{2-2}	-0.0004	-0.0004	0.0001
γ_{3-1}	-0.0006	-0.0006	0.0002
γ_{3-2}	-0.0000	-0.0001	0.0002
α_1	-0.1123	-0.1171	0.0193
α_2	-0.0038	-0.0026	0.0010
β_2	-0.8921	-0.9271	0.1472
Σ_{11}	0.0301	0.0298	0.0022
Σ_{12}	0.0024	0.0010	0.0002
Σ_{22}	0.0177	0.0154	0.0018
Date 1	100	100.09	1.3304
Date 2	200	202.71	5.5925

Table 4.3: DGP3 - Estimation Results When $m = 2$

Parameter	True values	MC mean	MC SD
μ_{1-1}	-0.0232	-0.0183	0.0139
μ_{1-2}	-0.0870	-0.0852	0.0293
μ_{2-1}	0.0725	0.0583	0.0128
μ_{2-2}	-0.0723	-0.0875	0.0532
μ_{3-1}	0.0058	0.0083	0.0037
μ_{3-2}	-0.0029	-0.0423	0.0144
γ_{1-1}	0.0029	0.0029	0.0009
γ_{1-2}	0.0017	0.0013	0.0008
γ_{2-1}	-0.0019	-0.0011	0.0013
γ_{2-2}	-0.0001	-0.0001	0.0000
γ_{3-1}	-0.0008	-0.0004	0.0013
γ_{3-2}	-0.0002	-0.0001	0.0001
α_1	-0.1652	-0.1388	0.0241
α_2	-0.0338	-0.0245	0.0099
β_2	-0.8552	-1.039	0.2822
Σ_{1-11}	0.0411	0.0418	0.0053
Σ_{1-12}	0.0058	0.0004	0.0001
Σ_{1-22}	0.0153	0.0142	0.0019
Σ_{2-11}	0.0103	0.0124	0.0061
Σ_{2-12}	0.0071	0.0070	0.0029
Σ_{2-22}	0.0087	0.0085	0.0037
Σ_{3-11}	0.0022	0.0033	0.0015
Σ_{3-12}	0.0009	0.0007	0.0003
Σ_{3-22}	0.0012	0.0009	0.0002
Date 1	100	99.558	2.1151
Date 2	200	198.69	6.0121

4.5 Application: Japanese Term Structure of Interest Rates

In this section, we analyse the Japanese term structure of interest rates using the cointegration model with multiple structural breaks presented in Section 4.2.

4.5.1 The Expectations Hypothesis

The term structure of interest rates states that the expected future spot rate is equal to the future rate plus a time-invariant term premium. For an overview of the expectations hypothesis theory, see Shiller (1990). The continuously compounded yield to maturity for an n period bond is defined as $r_{n,t} = -(1/n)\, p_{n,t}$ where $p_{n,t}$ denote the log of the price of a unit-par-value discount bond at date t with n periods to maturity, and the one-period future rate of return, earned from period $t+n$ to $t+n+1$, is given by $1 + F_{n,t} = P_{n,t}/P_{n+1,t}$. Let $y_{n,t}$ denote the yield to maturity n at t, then the expectations hypothesis implies:

$$y_{n,t} - y_{1,t} = n^{-1} \sum_{j=1}^{n-1} \sum_{i=1}^{j} E_t \left(\Delta y_{1,t+i} \right) + L_n \tag{4.27}$$

where $L_n = n^{-1} \sum_{j=0}^{n-1} \Lambda_j$ and Λ_j is the term premium. If $y_{1,t}$ is integrated of order one, then $y_{n,t}$ must be integrated of order one and $y_{n,t}$ and $y_{1,t}$ are cointegrated with cointegration vector (1, -1), that is analysed by Campbell and Shiller (1987). This cointegration relationship should be held in any pair of yield to maturity.

4.5.2 Estimation Results

We analyse Japanese term structure of interest rates for detecting structural breaks in a vector error correction model applying the method outlined in section 4.2. The data we use are 3-month bill rate and 1-year government bond yield based on the monthly data from IMF's *International Financial Statistics* ranged from June 1982 to April 2003 with 251 observations. These series are plotted in Figure 4.3. Figure 4.4 presents the spread between the two rates. Visually, there appears to be at least one structural break with potential break dates around 1988, 1990 or 1996, and the volatility of the two rates appear smaller after the 1996. The first differences of the two interest rates are shown in Figure 4.5, and it looks stationary throughout the entire observations and the volatility of the two rates seems to be decreased after late 1990's. To determine the number of structural breaks, we consider the VECM with $Y = \begin{pmatrix} \Delta r_s & \Delta r_l \end{pmatrix}$, where r_s denotes the short-term interest rate and r_l denotes the long-term interest rate, and estimate six models with structural breaks in the level, trend and covariance with $m = 0, 1, \ldots 5$. The number of lags in VAR is 12 selected by AIC and SBC. We adopt the same prior parameters used in the Monte Carlo simulations of the previous section. The Gibbs sampling is performed with 8,000 draws and the first 1,000 discarded.

The posterior probabilities for each model are $Pr\left(M_0 \mid Y\right) = 0.000\%$, $Pr\left(M_1 \mid Y\right) = 0.090\%$, $Pr\left(M_2 \mid Y\right) = 0.048\%$, $Pr\left(M_3 \mid Y\right) = 0.862\%$ and $Pr\left(M_4 \mid Y\right) = 0.000\%$ where the subscript of M denotes the number of breaks. Clearly, the no-structural break model is rejected by the data, and

$m = 3$ is strongly favoured.

The estimates of the parameters excluding the 11 lag terms of the vector error correction model with three structural breaks are given in Table 4.4. The results show that there are significant changes in the level, time trend and covariance. For example, the error covariance in the first regime is the largest and then becomes smaller as both rates approach to zero. This is not surprising since higher interest rates tend to fluctuate much more than lower rates.

The posterior means for the breaks are 1987.4, 1990.5 and 1995.7. The Japanese economy experienced an unprecedented bubbled expansion in late 1980s along with low interest rates. After 1987 the Bank of Japan increased interest rates causing the bubble to burst around 1990. After the collapse of the bubble, the Bank of Japan lowered the interest rates to recover the economic growth, and set the rates near zero after 1996.

The 95 % HPDR (Highest Posterior Density Region) for the break dates are also reported in Table 4.4, and the posterior probability mass function for the dates are shown in Figure 4.6. The distributions for the first and the third estimated dates cover relatively large intervals, while the second estimated date shows small range in distribution. If the range of the distribution is large, it might be possible to interpret that the break occurs smoothly over a number of consecutive dates smoothly. However, in our case of the second break date, as shown in the Figure 4.6, the break occurs in several neighborhood points not smoothly but discretely.

Figure 4.7 shows the posterior probability for the cointegrating vector, $\beta = (1, \beta_2)$, and the adjustment term, α. Note that $\alpha_i = 0$, $i = 1, 2$, is not included in the posterior densities. The expectation hypothesis tells that $\beta_2 = -1$ and this value is in the posterior density. More formal testing for this over-identifying restrictions on the cointegrating vector can be done by computing Bayes factor with the null of $\beta_2 = -1$ against the alternative of $\beta_2 \neq -1$. The Bayes factor is computed using (4.23) and (4.25) as $BF = \exp\left[0.5\left(\mathrm{BIC}_{UR} - \mathrm{BIC}_{R}\right)\right]$, where BIC_{UR} denotes the unrestricted BIC and BIC_{R} denotes the restricted BIC with the restrictions of $\beta_2 = -1$. The Bayes factor is 38.989, which shows strong evidence to support the expectation hypothesis according to Table 2.1.

Figure 4.3: Japanese long-term and short-term interest rates

solid line - short-term interest rate, dotted line - long-term interest rate

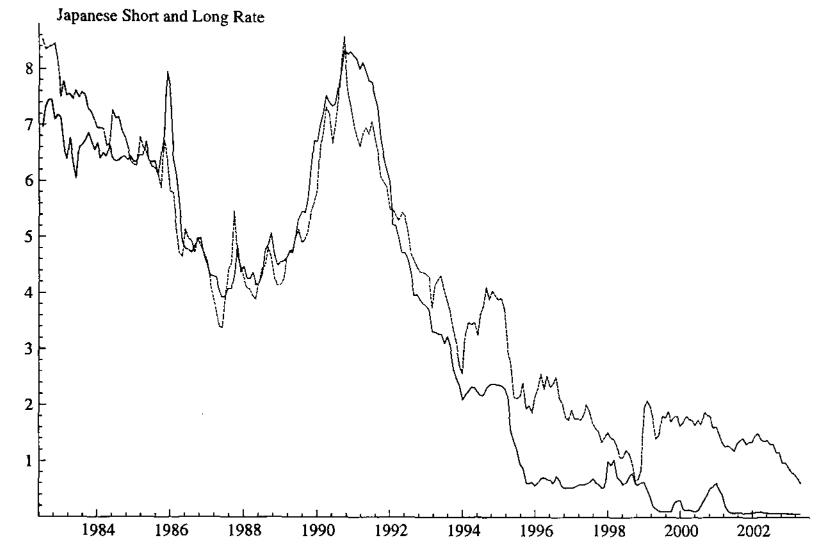


Figure 4.4: Spread between Long- and Short-term Interest Rates

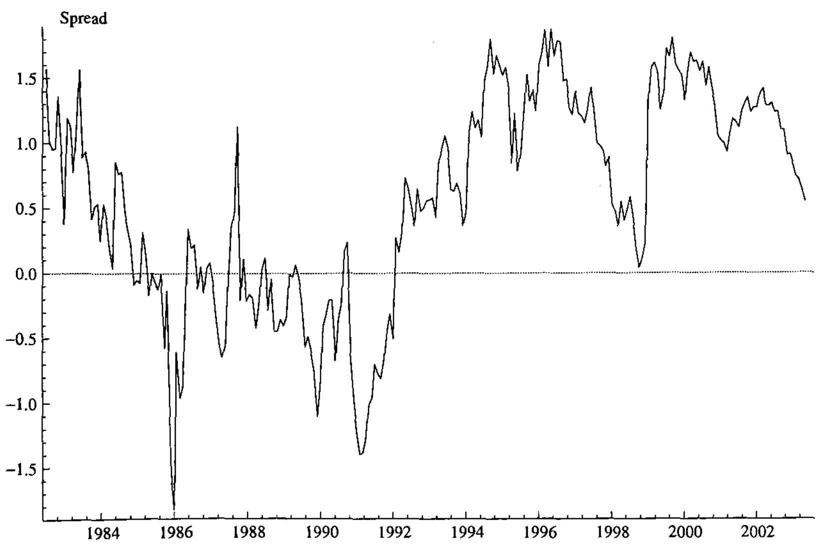


Figure 4.5: First Differences of Japanese Long- and Short-Term Interest Rates

solid line - short-term interest rate, dotted line - long-term interest rate

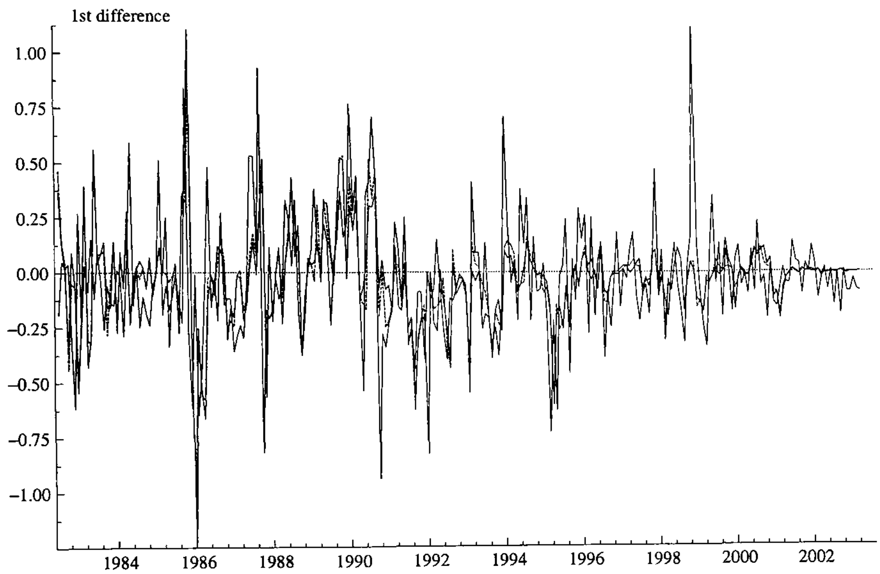


Table 4.4: Parameter Estimates for Japanese Term Structure

Model with 3 breaks in level, trend and covariance, ()=standard deviation,

Parameter	Posterior Mean	Posterior Mean
μ_1	0.002392 (0.01545)	-0.05777 (0.008099)
μ_2	0.1350 (0.02033)	0.06282 (0.01179)
μ_3	-0.09448 (0.009426)	-0.05359 (0.006454)
μ_4	-0.02321 (0.003359)	0.006943 (0.005457)
γ_1	-0.0004759 (4.054e-05)	-4.9899e-05 (6.291e-05)
γ_2	0.07856 (0.005229)	0.01787 (0.005457)
γ_3	0.2497 (0.006928)	0.05845 (0.007012)
γ_4	0.03918 (0.004389)	0.01945 (0.005331)
α	-0.03830 (0.005139)	0.04114 (0.007113)
β_1	1	-0.9966 (0.03563)

Table 4.5: Parameter Estimates for Japanese Term Structure

$$\Sigma_1 = \begin{bmatrix} 0.05611 & 0.02166 \\ 0.02166 & 0.1044 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.02890 & 0.01879 \\ 0.01879 & 0.04538 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 0.01390 & 0.01688 \\ 0.01688 & 0.06172 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 0.009874 & 0.002371 \\ 0.002371 & 0.02435 \end{bmatrix}$$

Estimates of the Break Dates

	Posterior Mean	95% HPDR
d_1	1987.3	1986.11, 1987.9
d_2	1990.4	1990.4, 1990.7
d_3	1995.7	1995.1, 1996.2

Figure 4.6: Posterior Probability Mass of the Break Dates

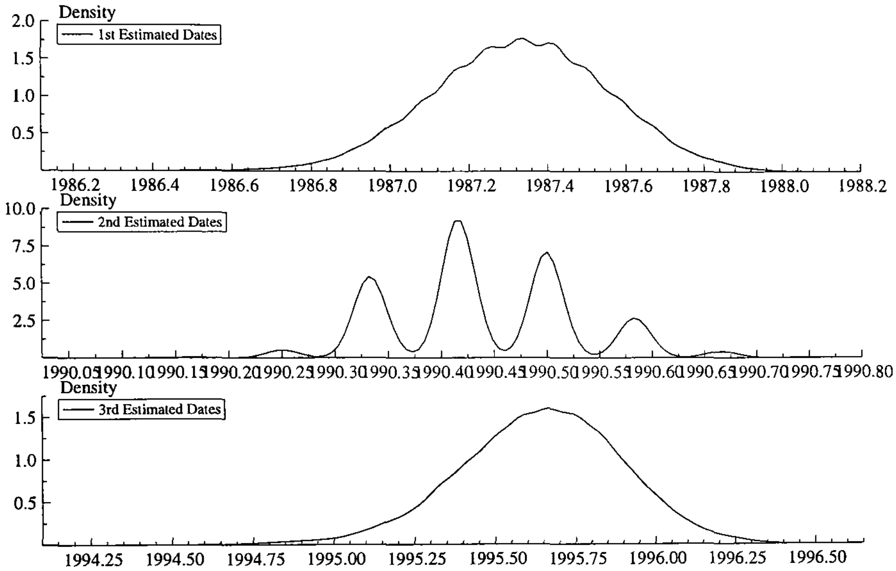
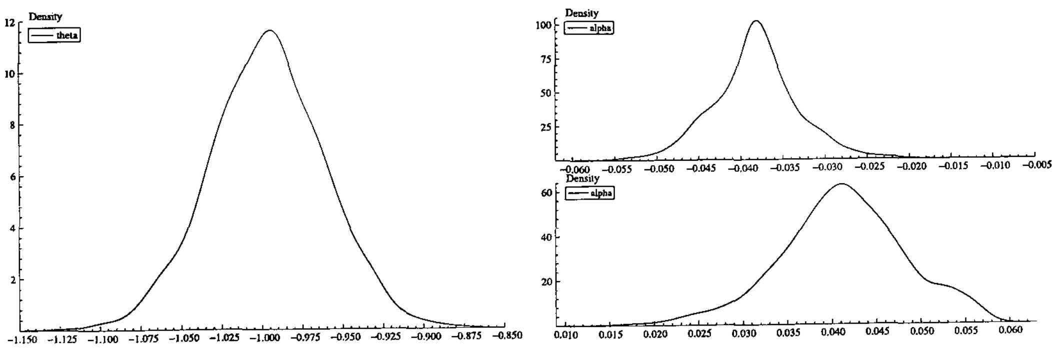


Figure 4.7: Posterior Density of β_2(left) and α (right)



4.6 Conclusion

We developed a Bayesian approach for analysing a dynamic co-integrated time series model with multiple structural breaks in the level, trend and error covariance based on the Gibbs sampler, extending Wang and Zivot's (2000) approach for univariate models. The number of break dates are chosen by the posterior probability based on the estimation of the model given the number of possible break dates. Our Monte Carlo simulations demonstrated that our approach provides generally accurate estimation for the number of break dates as well as their locations. Additionally, the Bayesian approach provides uncertainty in the location of the dates by the posterior mass function for each estimated break points.

In the application to the Japanese term structure of interest rates, we show that our Bayesian method is useful to analyse the case of multiple structural breaks. We found there is evidence of three structural breaks and the expectation hypothesis holds.

For future research, it is of interest that we compare our Bayesian methods with the classical methods of Bai, Lumsdaine, and Stock (1998). In this chapter we also present a more general cointegration model with multiple structural breaks in the adjustment term and the cointegrating vectors, in addition to breaks in the level, trend and error covariance. It is worth exploring this general models for its performance in Monte Carlo simulation and for its usefulness in the applications.

Chapter 5

Conclusion

In general, this thesis was concerned with Bayesian econometrics applied to non-linear cointegration time series models. We compare three Bayesian testing methods for determining the cointegration rank using Monte Carlo simulations. This analysis is then extended to the non-linear cointegration models with Markov regime switching, and finally to cointegration model with multiple structural breaks. We find that the Bayesian methods used in this thesis provide very flexible and simple solutions for analysing non-linear cointegration models.

5.1 Main Findings

In Chapter 2, we compare the performance of three Bayesian methods to test for the cointegration rank - the PIC, the KP method and a proposed method that uses Bayes factors. The procedure of the PIC method is not completely Bayesian. From the Monte Carlo results, we find that the PIC selects the true rank more frequently than other two methods, but only when the sample size is large. When the sample size is small, the PIC is not a good criterion to select the rank. This is not surprising since the PIC relies on some asymptotic properties. In small samples, the PIC tends to choose a lower rank, despite the fact that the PIC attaches twice as great a penalty to the rank of the cointegrating matrix than it does the parameters associated with stationary regressors (see Chao and Phillips). The PIC chooses one particular model specification among other models under consideration, but does not provide any measure of uncertainty of the models. This is one of the main weaknesses of the PIC from the Bayesian point of view. In contrast, the posterior odds calculated in the KP method and the method proposed in the chapter provide a measure of the model uncertainty.

For the KP method, the Monte Carlo results reveal that the method with conjugate priors gives the Bayes factors, computed by the Savage-Dickey density ratios, are greatly influenced by the specification of the prior's hyperparameters. With diffuse prior, the KP method produces similar results to the PIC. Therefore, if we focus on the selecting the cointegration rank, the KP method with diffuse prior may be a better choice since it gives us not only the cointegration rank but also a measure of uncertainty.

We present the simple method to detect the cointegration rank by computing the Bayes factors directly for each rank, and it has two versions - one method uses the Schwarz BIC to approximate the Bayes factors, the other method uses the harmonic mean of the likelihood. The Monte Carlo results show that when the sample size is small, the method using the harmonic mean of the likelihood has better performance over the other method approximated by the Schwarz' BIC; however, when the sample size is large, the method approximated by the Schwarz' BIC shows slightly better performance. Unlike the KP method which is invariant with respect to ordering of the variables in the VAR, the methods proposed in the chapter are not invariant to the values of the Bayes factors, although they tend to generate reasonable results whatever ordering is chosen.

Chapter 3 extends the framework used in Chapter 2 to allow the long-run multiplier to change according to the first-order Markov switching process. The Markov switching cointegration model can be useful when the researcher suspects that the cointegrating relationship changes according to the regime. A more general model where the adjustment term depends on the $m-$regime is also presented. As shown in Chapter 3 the estimation of these models is simple using the Gibbs sampler, that generates the posterior of all parameters of the model jointly. To estimate these models by the classical method, one has to estimate the cointegrating vector assuming the model is linear before making inferences on the unobservable regime variables. As shown in the Monte Carlo simulations and in the illustrative application to Purchasing Power Parity (PPP) between UK and US, there can be substantial differences in the estimated cointegrating vectors if the assumed model is linear as opposed to the assumed model is non-linear. The Bayes factors, approximated by the Schwarz'BIC, are used for selecting a model whether linear or non-linear model. Since the Bayes factors provide comparisons with all the models under consideration, either nested or nonnested, it is possible to compare linear no-cointegration models with non-linear cointegration models.

In Chapter 4, we extend the cointegration model to investigate the effects of multiple structural breaks. We focus on structural breaks in the level, trend and error covariance, but also show that it is quite easy to modify the analysis of the effects of breaks in the adjustment term and the cointegrating vector. These models show how the Bayesian method is flexible enough to permit estimating the model with unknown structural breaks in any subset of the parameters. The method also gives useful results of the uncertainty in the locations of the structural breaks, which can be interpreted as the gradual changing, rather than sudden jump to the other state.

5.2 Future Research

We select the conjugate prior with linear restrictions on the cointegrating vectors in this thesis. However, as Strachan (2003) and Villani (2003) pointed out, this prior might not be valid. Thus it is of interest if we consider their priors for our linear or nonlinear cointegration models. It is, also, of interest that we include their methods in our Monte Carlo simulations to compare with other methods.

This thesis deals with the cointegration model with Markov switching process in Chapter 3. The models assume that the regime changes occur suddenly at one date, which may not be realistic in some applications. Rather, we may consider smooth transition processes, which assumes that the regime changes occur gradually with some span of the time. Since the Bayes factors provide comparisons with either nested or nonnested models, it is possible to compare the smooth transition model or other models such as a threshold model with a Markov switching model by the posterior odds.

In this thesis we consider the two regime-switching models separately in Chapter 3 and Chapter 4. However, it is of interest that the Markov regime-switching model is combined with multiple structural breaks model. Kim and Nelson (1999) proposed a model which combines two Markov-switching processes - one imposed a restriction of zero on the transition probability to capture structural break, the other without any restriction on the transition probabilities to capture the business cycles, using Bayesian approach. However, it is possible to combine the structural break model presented in Chapter 4 with the Markov switching model presented in Chapter 3, and analyse, for example, Japanese business cycles in multivariate systems with structural breaks.

Bibliography

[1] Abuaf, N., and Jorion, P., 1990, Purchasing Power Parity in the Long Run, *Journal of Finance*, 45, 157-174.

[2] Albert, J.H., and Chib, S., 1993, Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts, *Journal of Business and Economic Statistics*, 11, 1-15.

[3] Andrews, D.W.K, 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica*, 61, 821-856.

[4] Andrews, D.W.K, Ploberger, W., 1994, Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative, *Econometrica*, 62, 1383-1414.

[5] Bai, J., 1997, Likelihood Ratio Tests for Multiple Structural Change, *Journal of Econometrics*, 91, 299-323.

[6] Bai, J., Lumsdaine, R., and Stock, J.H., 1998, Testing for and Dating Common Breaks in Multivariate Time Series, *Review of Economic Studies*, 65, 395-432.

[7] Bai, J., and Perron, P., 1998, Testing for and Estimation of Multiple Structural Changes, *Econometrica,* 66, 817-858.

[8] Balke, N.S., and Fomby, T.B., 1997, Threshold Cointegration, *International Economic Review*, vol.38, No.3, 627-645.

[9] Banerjee, A., Dolado, J.K., Galbraith, J.W., and Hendry, D.F., 1993, *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

[10] Banerjee, A., Lumsdaine, R., and Stock, J.H., 1992, Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence, *Journal of Business and Economic Statistics,* 10, 271-287.

[11] Bauwens, L., and Giot, P., 1998, A Gibbs Sampling Approach to Cointegration, *Computational Statistics*, 13, 339-368.

[12] Bauwens, L., and Lubrano, M., 1996, Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems, in T.B. Fomby (ed.), *Advances in Econometrics,* vol. 11B, JAI press, Greenwich, CT.

[13] Bauwens, L., Lubrano, M., and Richard, J-F.,1999, *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.

[14] Bertola, G., and Caballero, R., 1990, Target Zones and Realignments, *CEPR Discussion Paper* 3986, Centre for Economic Policy Research, London.

[15] Brenner, R.J., and Kroner, K.F., 1995, Arbitrage and Cointegration and Testing the Unbiasedness Hypothesis in Financial Markets, *Journal of Financial and Quantitative Analysis,* 23-42.

[16] Campbell, J.Y., and Shiller, R.J., 1987, Cointegration and Tests of Present Value Models, *Journal of Political Economy,* 95, 1062-1088.

[17] Campos, J., Ericsson, N.R., and Hendry, D.F., 1996, cointegration Tests in the Presence of Structural Breaks, *Journal of Econometrics,* 70, 187-220.

[18] Carlin, B.P., and Chib, S, 1995, Bayesian Model Choice via Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society,* Series B, 57, No. 3, 473-484.

[19] Carter, C.K., and Kohn, P., 1994, On Gibbs Sampling for State Space Models, *Biometrika,* 81, 541-553.

[20] Chao.J., and Phillps, P.C.B., 1999, Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure, *Journal of Econometrics,* 91, 227-271.

[21] Chen, J., and Gupta, A.K., 1997, Testing and Locating Variance Change-points With Application to Stock Prices, *Journal of the American Statistical Association,* 92, 739-747.

[22] Chen, M-H., 1994, Importance-Weighted Marginal Bayesian Posterior Density Estimation, *Journal of the American Statistical Association,* 89, 818-824.

[23] Chen, M-H., Shao, Q-M., and Ibrahim, J-G., 2000, *Monte Carlo Methods in Bayesian Computation,* Springer Series in Statistics, Springer.

[24] Chib, S, 1995, Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association*, vol.90, 432, Theory and Methods, 1313-1321.

[25] Chib, S, 1998, Estimation and Comparison of Multiple Change-Point Models, *Journal of Econometrics*, 86, 221-241.

[26] Chib, S., and Greenberg, E., 1995, Understanding the Metropolis-Hastings Algorithm, *The American Statistician,* 49, 327-335.

[27] Cristiano, L.J., 1992, Searching for a Break in GNP, *Journal of Business and Economic Statistics,* 10, 237-256.

[28] Clements, M.P., and Galvao, A.B, 2001, An Evaluation of Non-Linear Cointegrated Systems of the US Term-Structure of interest Rates, Unpublished Paper.

[29] DeJong, D., 1992, Co-integration and Trend-stationarity in Macroeconomic Time Series, *Journal of Econometrics*, 52, 347-370.

[30] Dickey, J., 1971, The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters, *The Annals of Mathematical Statistics* 42, 204-223.

[31] Doornik, J.A., 1998, *OX An Object-Oriented Matrix programming Language,* Timberlake Consultants Press., London.

[32] Dorfman, J., 1995, A Numerical Bayesian Test for Cointegration of AR Processes, *Journal of Econometrics*, 66, 289-324.

[33] Dreze, J.H., 1977, Bayesian Regression Analysis Using Poly-t Densities, *Journal of Econometrics*, 6, 329-354.

[34] Engel, C., 2000, Long-Run PPP May Not Hold After All, *Journal of International Economics,* April.

[35] Engle, R.F., and Granger, C.W.J, 1987, Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55, 251-276.

[36] Engel, C., and Hakkio, C.S., 1996, The Distribution of Exchange Rates in the EMS, *International Journal of Finance and Economics,* 1, 55-67.

[37] Evans, M. and Swartz, T., 2000, *Approximating Integrals via Monte Carlo and Deterministic Methods,* Oxford University Press, Oxford.

[38] Flood, R.P., and Garber, P.M., 1989, The Linkage Between Speculative Attack and Target Zone Models of Exchange Rates, *NBER Working Paper* 2918, Cambridge, MA.

[39] Frankel, J.A., 1986, International Capital Mobility and Crowding Out in the U.S. Economy: Important Integration of Financial Markets or Goods Markets?, In *How Open is the U.S. Economy?*, R.W.Hafer (ed.), Lexington Books, Lexington.

[40] Froot, K., Kim, M.C., and Rogoff, K., 1995, The Law of One Price Over 700 Years, *NBER Working Paper* 5132, Cambridge, MA.

[41] Froot, K., and Obstfeld, M., 1989, Exchange Rate Dynamics Under Stochastic Regime Shifts: A Unified Approach, *Harvard Institute of Economic Research Discussion Paper* 1451, Cambridge, MA.

[42] Garcia, R., 1998, Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, *International Economic Review*, 39, 763-788.

[43] Geweke, J., 1989, Bayesian inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, 57, 1317-1340.

[44] Geweke, J., 1996, Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics,*75, 127-146.

[45] Goldfeld, S., and Sichel, D., 1990, The Demand for Money, Chapter 8 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics*, Vol I, North-Holland, New York.

[46] Granger, C.W.J., 1981, Some Properties of Time Series Data and Their Use in Econometric Model Specifications, *Journal of Econometrics*, 16, 121-130.

[47] Granger, C.W.J., and Siklos, P.L., 1996, Temporary Cointegration with an Application to Interest Rate Parity, Discussion Paper 96-11, University of California, San Diego.

[48] Gregory, A.W., and Hansen, B.E., 1996a, Residuals Based Tests for Cointegration in Models with Regime Shifts, *Journal of Econometrics,* 70, 99-126.

[49] Gregory, A.W., and Hansen, B.E., 1996b, Practitioners Corner, Testing for Cointegration in Models with Regime and Trend Shifts, *Oxford Bulletin of Economics and Statistics,* 58, 3, 555-560.

[50] Hamilton, J., 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57(2), 357-384.

[51] Hansen, B.E., 1992, The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics,* 7, S61-S82.

[52] Hansen, B.E., 1996, Erratum: The Likelihood Ration Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics,* 11, 195-198.

[53] Hansen, B.E., 2000, Testing for Structural Change in Conditional Models, *Journal of Econometrics*, 97, 93-115.

[54] Hansen, P.R., 2003, Structural Changes in the Cointegrated Vector Autoregressive Model, *Journal of Econometrics*, 114, 261-295.

[55] Hansen, H., Johansen, S., 1999, Some Tests for Parameter Constancy in Cointegrated VAR-Models, *Econometrics Journal*, 2, 2306-333.

[56] Hastings, W.K., 1970, Monte Carlo Sampling Using Markov Chains and Their Applications, *Biometrika*, 57, 97-109.

[57] Hendry, D.F., 1995, *Dynamic Econometrics,* Oxford University Press, Oxford.

[58] Inclan, C., 1993, Detection of Multiple Structural Changes of Variance Using Posterior Odds, *Journal of Business and Economic Statistics*, 11, 289-300.

[59] Inclan, C., and Tiao, G.C., 1994, Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance, *Journal of the American Statistical Association*, 89, 913-923.

[60] Inoue, A., 1999, Test for Cointegrating Rank with a Trend-Break, *Journal of Econometrics*, 90, 215-237.

[61] Johansen, S., 1991, The Power Function for the Likelihood Ratio Test for Cointegration, in J. Gruber, ed., *Econometric Decision Models: New Methods of Modelling and Applications*, Springer Verlag, New York, NY, 323-335.

[62] Johansen, S., and Juselius, K., 1992, Testing Structural Hypothesis in a Multivariate Cointegration Analysis of the PPP and the UIP for UK, *Journal of Econometrics*, 53, 211-244.

[63] Kass, R.E., and Raftery, A. E., 1995, Bayes Factors, *Journal of the American Statistical Association*, 90, 773-795.

[64] Kass, R. E., and Wasserman, L. 1996, The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association*, 91, 1343-1370.

[65] Kim, C.J., and Engel, C., 1999, The Long Run U.S/U.K Real Exchange Rate, *Journal of Money, Credit and Banking*, 31, 335-356.

[66] Kim, C-J., and Nelson, C.R., 1998, Business Cycle Turning Points, A New Coincident Index and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching, *Review of Economics and Statistics*, 188-201.

[67] Kim, C-J., and Nelson, C.R., 1999, A Bayesian Approach to Testing for Markov Switching in Univariate and Dynamic Factor Models, mimeo.

[68] Kim, C-J., and Nelson, C.R., 1999, Has the U.S.Economy Become More Stable? A Bayesian Approach Based on a Markov Switching Model of Business Cycle, *Review of Economics and Statistics,* 81, 608-616.

[69] Kim, Y., 1990, Purchasing Power Parity in the Long Run: A Cointegration Approach, *Journal of Money, Credit and Banking,* 22, 491-503.

[70] King, R.G., Plosser, C.I., Stock, S.H., and Watson, M.W., 1991, Stochastic Trends and Economic Fluctuations, *American Economic Review,* 81, 819-840.

[71] Kleibergen, F., and van Dijk, H.K., 1994, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory,* 10, 514 - 551.

[72] Kleibergen, F., and Paap, R. 2002, Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration, *Journal of Econometrics*, 111, 223-249.

[73] Koop, G., 1991, Cointegration Tests in Present Value Relationships - A Bayesian Look at the Bivariate Properties of Stock Prices and Dividends, *Journal of Econometrics*, 49, 105-139.

[74] Koop, G., 1994, An Objective Bayesian Analysis of Common Stochastic Trends in International Stock Prices and Exchange Rates, *Journal of Empirical Finance*, 1, 343-364.

[75] Koop, G., 2003, *Bayesian Econometrics*, Wiley.

[76] Koop, G., and Potter, S.M., 1998, Bayes Factors and Nonlinearity: Evidence from Economic Time Series, *Journal of Econometrics*, 88, 251-281.

[77] Krugman, P.R., 1988, Target Zone and Exchange Rate Dynamics, *NBER Working Paper,* 2481, Cambridge, MA.

[78] Liu, J., Wang, W., and Kong, A. 1994, Covariance Structure of the Gibbs Sampler with Applications to the comparisons of Estimators and Augmentation Schemes, *Biometrica,* 81, 27-40.

[79] Lucas, R., 1988, Money Demand in the United States: A Quantitative Review, *Carnegie-Rochester Conference Series on Public Policy.* Vol 29, 137-168.

[80] Maddala, G.S., and Kim, I-M, 1998, *Unit Roots, Cointegration, and Structural Change,* Cambridge University Press, Cambridge.

[81] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N, Teller, A.H., Teller, E., 1953, Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 21, 1087-1092.

[82] Newton, M.A., and Raftery, A. E., 1994, Approximate Bayesian Inference by the Weighted Likelihood Bootstrap, *Journal of the Royal Statistical Society*, Ser. B, 56, 3-48.

[83] Nyblom, J., 1989, Testing for the Constancy of Parameters Over Time, *Journal of the American Statistical Association*, 84, 223-230.

[84] Paap, R., and van Dijk, H.K., 2003, Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to US Consumption and Income, *Journal of Business and Economic Statistics*, 21, 547-563.

[85] Perron, P., 1989, The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica,* 57, 1361-1401.

[86] Psaradakis, Z., Sola, M., Spagnolo, F., forthcoming, On Markov Error-Correction Models, With an Application to Stock Prices and Dividends, *Journal of Applied Econometrics.*

[87] Phillips, P.C.B. 1994a, Bayes Models and Forecasts of Australian Macroeconomic Time Series, in C. Hargreaves (ed.), *Non-Stationary Time Series Analysis and Cointegration*, Oxford University Press, 53-86.

[88] Phillips, P.C.B., 1994b, Model Determination and Macroeconomic Activity, Fisher-Schultz Lecture, European Meetings of the Econometric Society, Maastricht.

[89] Phillips, P.C.B., 1995, Bayesian Model Selection and Prediction with Empirical Applications, with comments by F.C. Palm and J.F. Richard and Reply by P.C.B. Phillips, *Journal of Econometrics*, 69, 289-365.

[90] Phillips, P.C.B. 1996, Econometric Model Determination, *Econometrica,* 64, 763-812.

[91] Phillips, P.C.B., and Hansen, B.E., 1990, Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies*, 57, 99-125.

[92] Phillips, P.C.B., and Loretan, M., 1990, Estimating Long-Run Economic Equilibria, *Review of Economic Studies*, 58, 407-436.

[93] Phillips, P.C.B., and Ploberger, W., 1994, Posterior Odds Testing for a Unit Root with Data-based Model Selection, *Econometric Theory,* 10, 774-808.

[94] Phillips, P.C.B., and Ploberger, W., 1996, An Asymptotic Theory of Bayesian Inference for Time Series, *Econometrica*, 64, 381-412.

[95] Poirier, D., 1995, Intermediate Statistics and Econometrics: A Comparative Approach, Cambridge, The MIT Press.

[96] Raftery, A.E., 1994, Changepoint and Change Curve Modeling in Stochastic Processes and Spatial Statistics, *Journal of Applied Statistical Science,* 7, 403-424.

[97] Ritter. C., and Tanner, M.A., 1992, Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association*, 87, 861-868.

[98] Saikkonen, P., 1991, Asymptotically Efficient Estimation of Cointegrating Regressions, *Econometric Theory*, 7, 1-21.

[99] Seo, B., 1990, Tests for Structural Change in Cointegrated Systems, *Econometric Theory*, 14, 222-259.

[100] Schwert, G.W., 1990, Stock Volatility and the Crash of '87, *Review of Financial Studies*, 3, 77-102

[101] Shiller, R.J., 1990, The Term Structure of Interest Rates. In: Friedman, B., Hahn, F. (eds), Handbook of Monetary Economics, vol. 1., North-Holland, Amsterdam, 627-722.

[102] Stock, J.H., and Watson, M.W., 1993, A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, *Econometrica*, 61, 783-820.

[103] Strachan, R., 2003, Valid Bayesian Estimation of the Cointegrating Error Correction Model, *Journal of Business and Economic Statistics*, 21, 185-195.

[104] Strachan, R., and Inder, B., 2004, Bayesian Analysis of the Error Correction Model, *Journal of Econometrics*, forthcoming.

[105] Strachan, R., and van Dijk, H., 2003a, Bayesian Model Selection for a Sharp Null and a Diffuse Alternative with Econometric Applications, Econometric Institute Report Series 2003-12, Erasmus University Rotterdam.

[106] Strachan, R., and van Dijk, H., 2003b, The Values of Structural Information in the VAR Model, Econometric Institute Report Series, 2003-17, Erasmus University Rotterdam.

[107] Tierney, L., and Kadane, J.B., 1986, Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association,* 81, 82-86.

[108] Toda, H.Y., and Phillips, P.C.B., 1994, Vector Autoregression and Causality: A Theoretical Overview and Simulation Study, *Econometric Reviews*, 13, 259-285.

[109] Tsurumi. H., and Wago, H., 1996, A Bayesian Analysis of Unit Root and Cointegration with an Application to a Yen-Dollar Exchange Rate Model, in T.B. Fomby (ed.), *Advances in Econometrics,* vol. 11B, JAI press, Greenwich, CT.

[110] Verdinelli, I., and Wasserman, L., 1995, Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American* Statistical *Association*, 90, 614-618.

[111] Villani, M., 2003, Bayes Estimators of the Cointegration Space, Sveriges Riksbank Working Paper Series No.150.

[112] Wang, J., and Zivot, E., 2000, A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance, *Journal of Business and Economic Statistics*18, 3, 374-386.

[113] Yao, Y-C. 1988, Estimating the Number of Changepoints via Schwarz Criterion, *Statistics & Probability Letters,* 6, 181-189.

[114] Zellner, A. 1971, An Introduction to Bayesian Inference in Econometrics, New York, John Wiley and Sons.

[115] Zellner, A. 1986, On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions, ed. P.K. Goel and A. Zellner, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti,* North-Holland, Amsterdam, 233-243.

[116] Zivot, E., and Andrews, D.W.K., 1992, Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, *Journal of Business and Economic Statistics,* 10, 251-270.