BAYESIAN INFERENCE IN NONLINEAR UNIVARIATE TIME SERIES: INVESTIGATIONS OF GSTUR AND SB MODELS

05D

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Fuyu Yang M.Sc. (Leicester 2004)

University of Leicester

Department of Economics

July 2008

Bayesian Inference in Nonlinear Univariate Time Series: Investigations of GSTUR and SB Models

Thesis submitted for the degree of
Doctor of Philosophy
by
Fuyu Yang
University of Leicester
Department of Economics
July 2008

Abstract
In the literature, many statistical models have been used to investigate the existence of a deterministic time trend, changing persistence and nonlinearity in macroeconomic and financial data. Good understanding of these properties in a univariate time series model is crucial when making forecasts. Forecasts are used in various ways, such as helping to control risks in financial institutions and to assist in setting monetary policies in central banks. Hence, evaluating the forecast capacities of statistical models, quantifying and reducing forecast uncertainties are the main concerns of forecast practitioners. In this thesis, we propose two flexible parametric models that allow for autoregressive parameters to be time varying. One is a novel Generalised Stochastic Unit Root (GSTUR) model and the other is a Stationary Bilinear (SB) model. Bayesian inference in these two models are developed using methods on the frontier of numerical analysis. Programs, including model estimation with Markov chain Monte Carlo (MCMC), model comparison with Bayes Factors, model forecasting and Forecast Model Averaging, are developed and made available to meet the demand of economic modelers. With an application to the S&P 500 series, we found strong evidences of a deterministic trend when we allow the persistence to change with time. By fitting the GSTUR model to monthly UK/US real exchange rate data, the Purchasing Power Parity (PPP) theory is revisited. Our findings of a changing persistence in the data suggest that the GSTUR model may reconcile the empirical findings of nonstationarity in real exchange rates with the PPP theory. The forecasting capacities of a group of nonlinear and linear models are evaluated with an application to UK inflation rates. We propose a GSTUR model to be applied with data, which contains as much information as possible, for forecasting near-term inflation rates.

ACKNOWLEDGEMENT

I wish to express my deepest thanks to my supervisors, Professor Stephen Hall, Professor Wojciech Charemza, and other scholars that also offered substantial help. In particular, I would like to thank Dr. Rodney Strachan, Dr Roberto Gonzalez and Professor Gary Koop. I am grateful to them for what I have learnt and I hope that I am able to live up to their very high standard.

I would like to thank the financial support from University of Leicester from 2004-2006 and also that from Economics Department from 2006-2008. Over my Ph.D. research, I am very fortunate to meet some top econometricians and economic theorists in conferences, seminars and workshops.

I would also like to express my gratitude to all my friends who have supported me over the last four years. Especially, I am grateful to Thomas Allen for his comments, and Simeon Coleman, Jonathan Chipili for their suggestions on my thesis, and Kavita Sirichand, Alessandro Tampieri, Jorge Villasenor for their friendship.

Many thanks to my parents for their support, encouragement and endless love throughout all my studies. Finally, this thesis would not be accomplished without a mention of Guoxia Bruce Yu, my husband. His continuing love and encouragement has given me the incentives to finish this thesis.

Contents

List of Tables

List of Figures

LIST OF NOTATIONS

Notation 1 $f_U(a,b)$ *denotes a uniform distribution on the interval of low bound a and up bound b.*

Notation 2 $f_{\exp}(x)$ *denotes an exponential density*

Notation 3 $f_N(\mu, V)$ *denotes a univariate normal density with a mean of* μ *and a variance of* V*.*

Notation 4 $\Phi_N(\cdot)$ *denotes the cumulative density corresponding to the standard normal density* $f_N(\cdot)$*, and the inverse cumulative distribution function is* $\Phi_N^{-1}(\cdot)$*.*

Notation 5 $f_{MN}(\mu, V)$ *denotes a multivariate normal distribution with a mean of* μ *and a variance covariance matrix of* V*.*

Notation 6 f_{χ^2} *denotes a Chi-square distribution.*

Notation 7 $f_{\chi^{-2}}$ *denotes an Inverse Chi-square distribution.*

Notation 8 f_Γ *denotes a Gamma distribution.*

Notation 9 f_Γ^{-1} *denotes an Inverse Gamma distribution.*

Notation 10 $1(A)$ *is an indicator function when event* A *happens.*

Notation 11 *In this thesis,* y *denotes the whole sample of observations with a sample size of* N*, where* $y = (y_1, y_2, \cdots, y_n)'$*. Initial value are treated based on the model specifications.*

Notation 12 $F_{t-1} = (y_1, y_2, \cdots, y_{t-1})'$ *denotes a sequence of past history up to* y_t*.*

Notation 13 *The rule of thumb in Bayesian framework: Posterior* $\propto$ *Likelihood* $\times$ *Prior*

Information for Users

The addenda to this thesis consisting of 1 CD.

Legal disclaimer: Although the author has done her best to ensure the programs in this CD are error free, she does not assume any responsibility for any errors which do exist. The author does not commit to provide user support for these programs. These programs may be freely used for educational and non-commercial research purposes.

Matlab R2007a is required for the programs. For the readers using a different software package (e.g. Gauss or Ox), the programs in this CD cannot be used directly. The CD consists of 6 folders in total, which are (1) toolkit for a Generalised Stochastic Unit Root (GSTUR) model, (2) toolkit for a Stationary Bilinear (SB) model, (3) toolkit for density forecasting with GSTUR, SB, Random Walk (RW) and stationary AR(p), and for combining forecast, (4) data sets applied in the thesis for empirical illustrations, (5) programs to plot probability density according to the distributions' properties, (6) toolkit for plotting.

Within each folder, there is a *readme* file in .txt format. It describes how the functions contained in the folder are executed, such as required input variables, return output variables, and algorithms applied within the functions.

In addition all of the programs begin with brief descriptions of what they do and are commented throughout with descriptions of what key commands do. As noted in the comments, many of these are taken from Jim LeSage's Econometrics Toolbox (www.spatial-econometrics.com) and some of the programs are adaptations of Gary Koop's codes (http://www.wiley.com/legacy/wileychi/koopbayesian/). Gary's book on introduction of Bayesian Econometrics (Bayesian Econometrics, 2003), which is accompanied with the programs, is highly recommended for Bayesian empirical applications.

Chapter 1

Introduction

This thesis has advanced nonlinear time series knowledge within a Bayesian statistical discipline. We seek to answer important questions with an intensive use of Markov Chain Monte Carlo (MCMC) and Bayesian econometric techniques, regarding the puzzles in economics of identifying a deterministic time trend, detecting nonlinearity in univariate time series, modelling complicated inflation dynamics, and quantifying forecasting uncertainties. On one hand, this thesis makes contributions to Bayesian econometric computation literature. On the other hand, by analyzing three macroeconomic time series, this thesis makes contributions to the literature of modelling stock price, real exchange rate series and inflation.

This thesis focuses on investigating two novel nonlinear models, a Generalised Stochastic Unit Root (GSTUR) and a Stationary Bilinear (SB) model, and providing Bayesian inference in these models using methods on the frontier of numerical analysis. A completed toolbox to calculate substantive quantities, such as model estimation, model marginal likelihood calculation and model-based point and density forecast is made available for empirical practice. The applied methodologies and procedure (programmed with Matlab R2007a) are extensively tested and validated on simulated artificial data, and then applied to real life macroeconomic data.

In section 1.1, we firstly motivate the nonlinear time series modelling approach with an application of Bayesian techniques. Section 1.2 summarizes the contributions of this piece of research and where it fits in the literature. Section 1.3 provides an abstract and an outline of each chapter.

1.1 Motivation

The main purpose of this thesis is to model the univariate macroeconomic time series dynamics using two novel nonlinear statistical models, which are the Generalised Stochastic Unit Root (GSTUR) model and the Stationary Bilinear (SB) model. In this section, to motivate our research, we seek answers for 3 questions.

(1). Why should we rely on numbers and statistical models to understand economic time series?

In Clive Granger's Noble Lecture (2003, Stockholm), he used metaphorical terms to describe a time series as "a loosely strung string of pearls". In his explanation, if the string is throw down onto a hard table top, a time series then can be represented by the string of pearls in a such way that "...the pearls give the points in the series", while "time is represented by the distance down the table". Therefore, "the placement of the pearl will impact where the next one lies because they are linked together".

Based on the ideas of "linked pearls", econometricians dedicate themselves to finding out the characteristics of the string that associate the places of the pearls, and predicting where the next pearl will be if the string extends. We may call the string an "underlying process" of the time series from a univariate time series perspective. Therefore, modelling underlying processes and forecasting based on historical data have an analogy to analyzing the movement of a string based on the placements of the pearls. Since it is commonly accepted that economic analysis should assist decision makers in making better decisions, a comprehensive understanding of the underlying processes and forecasting knowledge become essential in the univariate time series modelling. Given the facts that univariate time series modelling plays important roles in testing economic theories and providing forecasting, this thesis focuses on investigating 2 purely statistical models, the GSTUR and the SB models, for modelling univariate time series.

As well as investigating the "string", we look into the "pearls" as well. According to the frequency of collected data and sources from which we collect, different types of data have distinct characteristics. For example, the plots of high frequency data, such as stock prices collected every few seconds, are very different from those of low frequency macroeconomic data collected monthly. Analysis and modelling procedures of disaggregated data differ from those of aggregated data. For instance, the aggregated inflation rate is obtained via a weight aggregating approach to various price data from different sectors, where the weights can be time varying. In this thesis, we focus on low frequency macroeconomic data, which includes annual S&P 500 stock market

indices, quarterly UK inflation rates, and monthly UK/US real exchange rates.

(2). Why are nonlinear models appealing?

"The 1970s and 1980s saw economists adopt many of the time series techniques introduced by Box and Jenkins" (Potter, 1999). In recent years, there has been an increasing interest in the study of the nonlinear properties of macroeconomic and financial time series. In macroeconomics, the study of nonlinearity is based on an intuitive assumption that both the nature of the structural shocks hitting the economy and the dynamic properties of the economy might have changed. In financial applications, the introduction of nonlinearity is motivated by "theoretical models incorporating transaction costs" (Rapach and Wohar, 2006). Therefore, linear statistical models might be incapable of capturing the complicated dynamics in the underlying process. Davis and Resnick (1996) point out that "failure to account for nonlinearity can have dramatic consequences in the analysis and can be quite misleading".

In this section, we aim to provide a brief review of the expanding nonlinear time series modelling literature and some empirical findings regarding evaluations of nonlinear and linear models' forecasting capabilities.

The properties of many nonlinear statistical models have been investigated in the recent two decades since J. Hamilton's seminal paper on nonlinear modelling, in which a Markov Switching model was applied to US output, published in Econometrica 1989. Nonlinear models, such as Markov Switching and Markov Trend models (Koop and Potter, 1999), Random Coefficient Autoregressive model (Leybourne et al., 1996), Stochastic Unit Root (STUR) models (Granger and Swanson, 1997), Structural break model (Koop and Potter, 2004), Threshold autoregressive (TAR) model, Smooth Transition autoregressive (STAR) model, exponential smooth transition autoregressive (ESTAR) process, logistic smooth transition autoregressive (LSTAR) and their evolutive forms have been successfully applied in empirical macroeconomics and finance.

Regarding the forecast capacities of the statistical models proposed in the literature, it is not clear whether nonlinear models outperform linear ones, or *vice versa*. In Potter (1999), with a focus on modelling business cycles, he said that "successful nonlinear time series modelling would improve forecasts and produce richer notion of business cycle dynamics than linear time series allow". However, Potter caveats the statement with a reminder that the success of nonlinear forecasting models are based on two essential conditions: "First, economic time series must contain nonlinearity. Second, we need reliable statistical methods to summarize and understand these nonlinearities suitable for time series of the typical macroeconomic length". Hence, testing nonlinearity in a time series and identifying more precisely the type of nonlinearities

in the data is an important area for research.

Some researchers found that nonlinear models outperform linear models in applications to macroeconomics and finance. Clement and Smith (2000) compare nonlinear models (self-exciting threshold autoregressive models) and linear models by evaluating forecast densities with an application to output growth and unemployment. In their example, nonlinear models outperform the linear models, such that nonlinear models are "better able to predict higher order moments" (Clement and Smith, 2000). With applications to a panel of spot exchange rates and spot interest rates, Canova (1993) found a multivariate Bayesian time-varying coefficient (TVC) approach to forecasting improves over a Random Walk model. With applications to postwar U.S. quarterly inflation forecasting, Watson and Stock (2007) compared the forecasting model capabilities amongst multivariate models and univariate models. Models they considered are AR models with lag lengths selected by the AIC (AR-AIC), Random Walk (RW) model, rolling MA(1), backwards-looking Phillips curve (PC), a trend-cycle model and an unobserved component model with stochastic volatility (UC-SV). They propose a UC-SV model to forecast inflation rates because of its good overall out-of-sample forecast performance. Compared amongst linear and nonlinear models, an AR-AIC does not forecast as well as the UC-SV model. Compared amongst linear models, a rolling MA(1) improves upon other univariate linear forecasting models. Interestingly, they conclude that multivariate forecasts do not improve on forecasts made using time-varying univariate models, such as unobserved trend-cycle model with stochastic volatility or integrated moving average process with time-varying parameters.

With applications to different data sets and different nonlinear model specifications, Stock and Watson, however, did propose linear against some nonlinear models in their early research. Stock and Watson (1998) performed a comparison of linear and nonlinear univariate models for forecasting 215 US monthly macroeconomic time series at three forecasting horizons over the period 1959-1996. Models they considered are AR, ESTAR, artificial neural networks (ANN) and LSTAR. They found that AR have lower average cost than LSTAR and ANN. Simple forecasting AR models with lag lengths selected by the AIC are also supported by Meese and Geweke (1984), who compared various linear models using 150 macroeconomic time series.

Some researchers choose not to be biased between nonlinear and linear models. With an application of monthly real exchange rates from the post-Bretton Woods period, Rapach and Wohar (2006) found point forecasts generated from Band-TAR and ESTAR models are very similar to forecasts generated by linear AR models at short horizons. Their overall conclusion is that "nonlinearities (in the data period) are

too slight or subtle for the band-TAR and ESTAR model specifications".

Given the voluminous literature and ambiguous conclusions, we address 5 questions in the context of investigating nonlinearities in univariate time series. First, do nonlinearities exist in univariate time series and if so, what is the type of the nonlinearities? Second, do nonlinear time series models produce out of sample forecasts that improve upon linear models? In particular, do nonlinear models outperform Random Walk models in the context of forecasting the inflation rate? Third, can the unobservable changing persistence be detected in a univariate time series with applications of flexible nonlinear models? Fourth, to what extent are we sure that a deterministic trend exists in the series? Fifth, if the underlying process is a quasi-Random Walk[1] process, can we trust the nonstationarity test results from conventional Augmented Dickey-Fuller (ADF) test and/or Phillips-Perron test? To avoid raising topics that are too broad to cover, we restrict ourselves to 3 specific univariate time series: S&P 500 annual data from an extended Nelson and Plosser's data set (from 1877 to 1988), UK/US long run monthly real exchange rates from January 1885 to February 1995 with a sample size of $1,322$ over 111-year period, and the UK quarterly inflation rates from 1957 Q1-2007Q1. Also, we consider nonlinearities in two specific forms: the GSTUR process and SB process.

(3).Why Bayesian?

This section introduces our motivation of Bayesian approaches to time series modelling.

Bayesian techniques has been widely applied in time series modelling, such the revisits of Nelson and Plosser's data and finding the evidence in support of deterministic time trends in DeJong and Whiteman (1991a, 1991b)[2] and also in Kwiatkowski et al. (1992), analysis of unit root in 8 real exchange rate series and finding empirical evidence of PPP in French franc/German mark in Schotman and van Dijk (1991), and finding evidence of nonlinearities in US GNP in Koop and Potter (1999).

The advances in simulation-based technology has absolutely revolutionized Bayesian approaches to time series modelling (Geweke, 1989). One of the current research strands in applied Bayesian statistics is to develop Markov chain Monte Carlo (MCMC) methods for complicated time series models. Examples can be found in Chib et al. (2002) and Chib et al. (2006), where sampling algorithms are developed to estimate Stochastic Volatility (SV) model and multivariate SV. Also, Bauwens and Lubrano (1998) applied Gibbs sampler to estimate GARCH model in Bayesian framework.

[1] Also, we can call a quasi-Random Walk process as a Random Walk like process.

[2] Or "trend-stationarity" is the terminology they used.

In Koop et al. (1997), the advantages of the Bayesian approach to time series modelling over classical techniques are summarized as the following: (1) It provides exact finite sample distributions for any feature of interest. (2) Instead of presenting just a point estimate associated with a standard error, the whole density of the quantity of interest can be plotted. (3) Comparisons amongst models specified in any form can be achieved. (4) Forecast uncertainties can be quantified where model uncertainty can be accounted for. Using MCMC, density forecast can be achieved without extra cost.

Feature (1) is very appealing, especially when estimations of latent variables are involved. Classical frequentist confront estimation problems when the entertained models are high dimension. However, with the use of the MCMC method, any model is tractable with Bayesian approaches. Interpreting the summarized posterior statistics, such as Bayesian credible intervals, is straight forward and intuitive. Feature (2), presenting the whole density is important since sometimes the parameter of interest could be multi-modal, skewed and fat-tailed. Appropriate presentations of these properties are crucial in forecasting, in particular controlling forecast risks when parameter uncertainties are accounted for. Feature (3) is related to model selections. Unlike the classical method, the Bayesian approach to testing does not rely on asymptotic or approximations. Finally, feature (4) relates to the Bayesian forecasting and Bayesian model averaging (BMA) provides straight forward solutions to combining forecasts and reducing forecast uncertainties.

In time series modelling, Bayesian and classical methods gives different answers to unit root tests. However, even amongst Bayesian econometricians, agreement is not reached with respect to prior selections and testing procedures. In the 1990s, Nelson and Plosser's macroeconomic data is tested in a Bayesian framework by Sims and Uhlig (1991), Phillips (1991), DeJong and Whiteman (1991a, 1991b), Schotman and van Dijk (1991), Koop and Steel (1991). Some researchers, such as Kim and Maddala (1991), try to tackle the disagreements focusing on the prior selections in AR(1) model with Monte Carlo simulations. Bauwens et al. (1999) summarize this series of unit root debate and address the following caveats when we apply Bayesian approaches to modelling time series:

1. Construction of priors is important, because different priors not only represent different prior belief, but also bring about different properties into the posteriors via the likelihood function. For illustrations, see Bauwens et al. (1999, pp.174-192), where the Jeffreys' prior, Phillips' prior, Berger and Yang's (1994) prior and Lubrano's prior are evaluated in the context of a simple AR(1) model.

2. Specifications of deterministic terms matters. Excluding a deterministic time trend may result in different inference in the unit roots. See the results in DeJong and Whiteman (1991a).

3. The initial conditions do not matter in a large sample for a stationary processes, but do matter in the case of a nonstationary process (Bauwens et al., 1999, pp.164).

4. When we do testing, or more precisely, model comparison, how the models are defined matters. See examples in Schotman and van Dijk (1991).

In chapter 2, we provide a review in details with respect to comparisons between the classical approach and Bayesian approach to model estimation, model selection and forecasting.

Bauwens et al. (1999, Chapter 6) provides intensive illustrations on how to apply Bayesian techniques with various dynamic econometric models, such as an AR(p) model in chapter 6, Autoregressive Conditional Heteroscedasticity (ARCH) model in chapter 7. In particular, chapter 8 provides comprehensive Bayesian inference with respect to nonlinear time series statistical models, which are the Threshold autoregressive (TAR) model and its extensions. For empirical illustrations, they applied a logistic smooth transition autoregressive (LSTAR) model to analyze the asymmetries in the US business cycles. Interested readers can also refer to Koop (2007, chap. 17 and 18), which provides details of making Bayesian inference in statistical time series models, such as stationary AR(p) model, basic TAR and TAR with switches in the Error Variance, and ARCH model.

1.2 Main Contributions

This thesis focuses on investigating two novel nonlinear models, and providing Bayesian inference in these models using methods on the frontier of numerical analysis. With respect to these nonlinear models, a completed toolbox to calculate substantive quantities, such as model estimation, model marginal likelihood calculation and model-based point and density forecasts, is made available for empirical practice. The applied methodologies and procedures (programmed with Matlab R2007a) are extensively tested[3] and validated on simulated artificial data, and then applied to real life macroeconomic data.

[3]Sampling efficiency is tested according to Carlin and Louis (2000, pp.182).

There are a number of main contributions of this thesis that add knowledge to univariate time series modelling from a Bayesian perspective. On one hand, this thesis makes contributions to Bayesian econometric computation literature. On the other hand, by re-visiting three univariate macroeconomic time series, which have been tested for economic theories by many times, this thesis makes contributions to the literature of nonlinearity, deterministic time trend and changing persistence in empirical macroeconomic modelling.

Contributions to Bayesian Econometric Computation:

In chapter 3, in order to handle more complex dependence in time series, *the first* contribution is that we extend a Stochastic Unit Root (STUR) model, which was originally proposed by Granger and Swanson (1997), to a Generalised Stochastic Unit Root (GSTUR) model. Within a Bayesian framework, we apply Markov chain Monte Carlo (MCMC) methods and developed an efficient[4] posterior simulator for this high dimensional GSTUR model. We solve the problem of estimating high dimensional latent variables, which cannot be resolved satisfactorily in a classical framework.

On top of the contribution of providing an efficient algorithm to estimate the GSTUR model, we also provide an algorithm to evaluate the marginal likelihood with a GSTUR specification following the idea in Chib, Nardari and Shephard (2002). Because analytical methods for marginal likelihood evaluations are not applicable in the GSTUR model, the algorithm we developed overcomes the problem of integrating high dimensional latent variables. This algorithm utilizes a method proposed by Chib (J. Amer. Statist. Assoc. 90 (1995) 1313) and facilitates an Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (J. Amer. Statist. Assoc. 94 (1999) 590), where the advance makes the GSTUR model available to be compared with any model of interest.

The second contribution is made by providing a complete simulation-based statistical analyzing toolkit with a focus on a Stationary Bilinear (SB) model within a Bayesian framework in chapter 4. With respect to the nonlinear time series modelling discipline, this toolkit makes the SB model easily available for applications.

We propose that the SB model be estimated using the MCMC algorithms we developed. Validated on simulated data, estimates obtained via Bayesian MCMC with a loose prior are as equally efficient as the estimates obtained via the Classical maximum likelihood (ML). With a tight prior, the MCMC provides more accurate

[4]In this thesis, we used "efficiency" and "efficient" in many places. To clarify: efficiency of an estimator indicates small variance. Computational efficiency of an algorithm means short execution time, and efficiency of a simulator means it approximates the target distribution well and yield estimates of posterior quantities with high RNEs.

estimates than the ML estimates. For MLE approach to estimate the SB model, refer to Hristova (2005). Also, we developed an algorithm to evaluate the marginal likelihood of SB, which utilizes a Gelfand-Dey method. This algorithm for marginal likelihood calculation allows the SB to be compared with any model of interest, such as the GSTUR model we proposed in chapter 3, and a linear Random Walk (RW) model.

An attempt is also made to distinguish a Random Walk like process, the SB process, from a Random Walk process using Bayes Factors within a Bayesian framework. We re-examined the testing power of the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests, using a data series simulated from the SB data generating mechanism. We found that the conventional ADF and PP tests perform poorly against the SB process, providing misleading results of a unit root.

The third contribution is that we add knowledge to the nonlinear model-based forecast literature in chapter 5. Two strands in the literature, which are forecasting with nonlinear models in a Bayesian framework and forecast combinations, are brought together.

In chapter 5, the idea of calculating substantive quantities with the GSTUR and SB models, such as model-based forecasting, is further developed. The forecast capacities[5] of 24 statistical forecasting models are evaluated. The 24 models we considered include the GSTUR class models, SB model, linear stationary AR models with lag length selected according to AIC, and a linear RW model. Results show that the GSTUR forecasting model performs the best with an application of the quarterly UK inflation rates.

Regarding the issue of combining forecasts, we revisit the puzzle in the forecast combination literature and re-raise the question: Does a combined forecast from a Simple Averaging (SA) approach improve upon those forecasts from individual forecasting models, and improve upon a combined forecast from a Bayesian Model Averaging (BMA) approach? Our results is consistent with the argument in Meese and Geweke (1984), in a sense that SA does not always improve upon the forecasts from individual forecasting models. A difference between our research and the work in Meese and Geweke (1984) is that constituent models they considered are linear models, with a Random Walk as the benchmark model. In our case, the model space contains 21 nonlinear models.

One limitation in chapter 5 is that only 24 models are consider. In applications with real settings, it is possible that some complex dynamics may not be fully cap-

[5]The forecast capacity is referred as both the point forecast and density forecast capacities.

tured, by either any of the constituent forecasting models, or combined forecast, or neither. Therefore, in future research we may enlarge the model space by taking other models, such as TAR, Markov Switching, Stochastic Volatility models, into consideration. Another limitation is that the one-step ahead out-of-sample forecast is carried out for 10 periods only. Therefore, the density forecast evaluation results may be distorted due to the low testing power against small forecast samples.

Contributions to Discrete Low Frequency Time Series Modelling:

With applications of the toolkits we developed for GSTUR and SB models, we made 4 main contributions by adding knowledge to discrete low frequency macroeconomic time series modelling.

First, the theory of Purchasing Power Parity (PPP) is tested using the UK/US monthly long run real exchange rates. We propose a GSTUR model for reconciling the nonstationarity in macroeconomic time series with the PPP theory. We found persistence has shifts within the sample, along with changes in the monetary events. Using the same data series, Engel and Kim (1999) applied a Markov Switching model and identified a permanent component and a transitory component. What they found is that "the transitory component appears to be driven by temporary monetary phenomena, and the shifts of variance occur at times of historically important monetary events". Our research tries to model the real exchange rate data from another perspective by looking into the changing persistence, which appears to be correlated with significant monetary events.

One limitation is that we conclude the correlation between the changing persistence and the significant monetary events via visualizing a plot. We did not investigate to what extent the changing persistence and monetary events are correlated.

Second, the existence of a deterministic time trend in an extended Nelson and Plosser's data is re-investigated. We address the following questions: how sure are we that economic time series have deterministic trends if the underlying properties have changed? With an application of S&P 500 annual data to the GSTUR class models, we found strong evidence of a deterministic trend and changing persistence according to the model probabilities of the GSTUR class of models specified with deterministic time trends. The result of supporting a deterministic trend is consistent with the findings in DeJong and Whiteman (1991a, 1991b). However, given the fact that GSTUR process cannot be differenced to stationarity, we may conclude that the S&P 500 series we analyzed is neither trend-stationary series nor a difference stationary process.

One limitation in this empirical application with S&P 500 series is that we did not evaluate the forecasting capacity of a model specified as a GSTUR with a determin-

istic trend. Further investigations are needed and the results can be fitted into the forecasting literature. Hendry and Clements (2003) claim that difference stationary models produce better forecasts when compared with trend-stationary models. An interesting question can be asked: if the underlying properties have changed, will the models excluding a deterministic time trend still produce better forecasts?

A third contribution is made in chapter 4. With an application of the quarterly UK inflation data set, we found that the inflation rate can be better represented by the SB process compared with a RW process.

A final contribution is adding knowledge to the literature of forecasting UK inflation rates. By fitting a GSTUR model to the quarterly UK inflation data, we found the estimated stochastic roots follow entirely different patterns before and after the 1990s. The evidence of changing persistence in inflation can also be found in a very recent working paper by Cogley, Primiceri and Sargent (2008) using the US data[6]. With respect to forecasting, our results show that the GSTUR model specified with a constant represent the best forecast performance in both full sample and small sample.

In previous research, Atkeson and Ohanian (2001) claim that Phillips-curve based inflation forecasting models cannot do better than a Random Walk. According to our results that GSTUR provides better forecast than the RW, which agrees with Watson and Stock (2007) on the point that inflation dynamics should be modelled with time-varying parameter models. A limitation in this regard is that the GSTUR model we proposed considers changing persistence. However, trend-cyclical components and stochastic volatility are omitted. More extensions are left for future research and whether a GSTUR model encompasses other nonlinear models, such as Markov Switch model or Structure break model, is open to research.

1.3 Abstract and Outline of Chapters

In order to provide a high level overview of Bayesian solutions to the questions we would like to address in chapter 1.1, Chapter 2 provides a general background of the Bayesian framework and some computational techniques carried through in this thesis. The chapter consists of five main parts. Firstly, we compare the frequentist and Bayesian theories in a high level of abstraction, by reviewing their different approaches to estimation and hypothesis testing. Secondly, since Bayesian methods are criticized as being 'subjective' with respect to the prior eliciting procedure, we provide

[6]More precisely, they evaluated the persistence of inflation gap, which is defined as "the deviation of inflation from a pure random walk component of inflation".

the Bayesian viewpoints in response to this criticism. Thirdly, we elaborate a commonly applied Bayesian estimation method, the Monte Carlo Markov Chain (MCMC) strategy, and its advantages, included algorithms, as well as its diagnostic methods. Fourthly, we review the Bayesian literature of developed computational techniques in terms of model assessment and selection. Finally, since the applicabilities of the models of interest are our main concerns, we focus on the reviews of Bayesian forecasting and Bayesian Model Averaging (BMA) in the literature.

Chapter 3 makes use of the novel Generalised Stochastic Unit Root model to investigate the presence of a deterministic time trend in economic series. The model is specified to allow for changes in persistence over time, such as shifts from stationarity (I(0)) to nonstationarity (I(1)) or vice versa. This paper proposes a Generalised STUR model for reconciling the nonstationarity in macroeconomic time series with the economic theory, such as the Purchasing Power Parity puzzles (see Taylor et al. 2001). The use of the Generalised STUR model is also motivated by the fact that there is always uncertainty as to whether a macroeconomic series (see Newbold et al. 2001) is trend stationary (TS) or difference stationary (DS) or neither. This uncertainty raises the crucial question about how sure one can be that an economic time series has a deterministic trend when there is a change in the underlying properties. In this paper, Bayesian model estimation and Bayesian model comparison techniques are utilized in an investigation of a deterministic time trend while the time series properties are relaxed from constant I(1) or constant I(0) to a stochastic unit root process. An application to the series of Standard & Poor 500 indices (S&P 500) indicates that in the S&P 500 series, it is plausible that a deterministic time trend exists together with time-varying persistence (in the form of STUR). A further application to UK/US long-run real exchange rate also indicates that the STUR model could provide new insights in time series analysis.

The outline of chapter 3 is as follows: Section 1 provides some puzzles in time series analysis and motivates the GSTUR model. Following the logic of Bayesian reasoning, Section 2 develops a posterior simulator to estimate the high dimensional GSTUR model, evaluates the marginal likelihood, compares amongst a class of GSTUR model along with a Random Walk (RW) model, and all the developed tools are investigated in controlled settings using an artificial data set. Armed with the tools developed in Section 2, Section 3 carries out empirical analysis using the two well known time series: S&P 500 indices from the extended Nelson and Plosser's data set (Nelson and Plosser, 1982), and the long run UK/US real exchange rates.

Chapter 4 takes on research of another nonlinear model. In recent years, there has

been an increasing interest in the study of the nonlinear and non-stationary properties of macroeconomic and financial time series. However, it is difficult to make conclusive decisions on the non-stationarity of a time series as most statistical tests have low power in distinguishing random-walk-like processes from the Random Walk (RW) processes. In this paper, Bayesian methods are applied to a stationary bilinear (SB) model, in which autoregressive (AR) coefficients are restricted within and close to the boundaries of stationarity, meanwhile the AR coefficients are correlated with the random error terms. A posterior simulator for model estimation, alongside a toolkit for model forecasting are developed. By adopting Bayesian approaches, non-nested model comparisons, amongst the nonlinear SB model and a group of linear models (RW model, AR (p) models), are easily achievable. With an application to quarterly UK inflation rates (1957Q1- 2007 Q1), we found that this stationary bilinear model could well represent the underlying process of the time series, whilst a unit root cannot be rejected using the conventional Augmented Dickey Fuller (ADF) test. According to the Bayes Factors (BF) from the model comparisons results, the SB model receives a higher model probability compared with a RW model.

Chapter 4 consists of two main parts. In the first part, Section 1, we motivate the Stationary Bilinear (SB) model. Then we provide Bayesian inference with respect to this statistical model in Section 2 and 3. In the second part, the developed tool pack is evaluated in Section 4 and applied with the UK's inflation rates in Section 5.

Chapter 5 investigates the UK inflation rates forecasting from a univariate time series perspective based on 5 sources (Clements and Hendry, 1998) that normally induce forecast uncertainties: (1) Forecast failure is often due to the inherent non-stationarity of the data, and future abrupt changes in the underlying system, which are unpredictable. With an application to the quarterly UK inflation rates data, we tried to tackle questions, such as whether nonlinear models are more resilient to structural breaks than linear models in terms of forecasting, via evaluating one-step-ahead out-of-sample point forecast performance of a pool of nonlinear and linear statistical forecasting models. Model spaces consists of the Generalised Stochastic Unit Root (GSTUR) class models, Stationary Bilinear (SB) model, Random Walk model and Stationary AR(p) model. We found that, the GSTUR models are better at predicting the point forecasts and providing smaller Mean Squared Forecast Errors (MSFE), compared with other entertained models in the model space. (2) Model uncertainty is another source of forecast uncertainty. Combining the forecasts from the constituent models is a common way to reduce the uncertainty in forecasting practice. This chapter applies a Simple Averaging (SA) and a Bayesian model averaging (BMA)

approach to combining forecast densities of the constituent forecasting models. (3) Mis-measurement of the data in the base period from which forecasting begins also causes forecast uncertainties. To answer the question, such as whether we should simply abandon partial historical data when a structural break is detected in the series, we forecast the same period from 2004Q4 to 2007Q1 based on the same forecasting models using two different historical samples. One is the full sample of data (Q1 1957-Q1 2007) with 201 observations and the other is a sub-sample of the data (Q4 1999-Q1 2007) with 30 observations after the independence of the Bank of England. Our finding is that although a structural break may induce bigger forecast errors, the lack of data will induce larger forecast uncertainties according to the inflated variance of simulated forecast distribution when the small sample is applied. (4) Forecast uncertainties are also due to the inaccuracies in the estimates of the model's parameters, and (5) the cumulation of future errors (or "shocks") to the economy. We constructed density forecasts of the proposed forecasting models via simulations to quantify the forecast uncertainties induced by sources (4) and (5). The density forecasts are evaluated with a Probability Integral Transform (PIT) approach. Finally, a tool kit, which includes constructing point and density forecasts of GSTUR, SB, RW and AR(p) model, combining forecasts via SA and BMA approaches, evaluating forecast results, graphic tools to present forecasts, is prepared for forecast practitioners.

Chapter 5 is organised as follows: section 1 provides a brief review of current forecasting puzzles we encountered in the economic discipline and a list of questions we would like to investigate in the chapter. Section 2 introduces the methodologies applied in chapter 5, which includes issues related to forecast constructions, combining forecast and forecast evaluations. Section 3 presents the statistical forecasting models in the model space and forecast algorithms. Section 4 applies the methodologies reviewed in section 3 to the UK inflation rate data. Section 5 concludes.

Finally, we summarize the main findings of the thesis in chapter 6. Also we make suggestions for future research.

Chapter 2

A Survey of Bayesian Techniques

This Chapter aims to provide a general background to the Bayesian framework, and to motivate the application of to the Bayesian method for statistical analysis. In Section 1, some fundamental concepts of the Bayesian framework are introduced and the differences between using the classical frequentist approach and the Bayesian approach to making statistical inference are briefly reviewed. Section 2, addresses the criticisms of Bayesian prior selections, and some solutions to the prior elicitations are surveyed. Section 3 provides the Bayesian solution to model estimations: the Markov Chain Monte Carlo (MCMC) method. The sampling techniques that we reviewed will be intensively applied in later chapters. Section 4 focuses on Bayesian solutions to model selections, and Section 5 reviews the perspectives of Bayesian forecasting and Bayesian Model Averaging.

2.1 Preliminaries

This section aims to provide and compare some fundamental issues regarding the frequentist inference approach and the Bayesian inference approach to statistical analysis. After a short introduction to the nature of the Bayesian approach, hypothesis testing and point estimations in both frameworks are reviewed.

The main difference between the frequentist and the Bayesian statistical philosophies is based on the understanding of probabilities. In the frequentist framework, probabilities are handled within well defined random repeated experiments and the probability of an event to occur depends on the limiting relative frequency in a large number of future hypothetical experiments. One of the main criticisms of the frequentist approach is based on non-repeatable experiments, in the sense that we may not be able to prove or refute the propositions we are holding.

In the Bayesian framework, probabilities are regarded as "degrees of belief " such that we can "associate numerical probabilities with degrees of confidence that we have in propositions about empirical phenomena" (Zellner (1971, pp.9)). Therefore, the uncertainties attached to an event can be quantified with some form of numerical standards and revised according to the available information. Also preferences between options can be compared. This quantification approach to uncertainties is described as "quantitative coherent", which is addressed by Bernardo and Smith (1994, pp.444) in terms of "... an axiomatic foundation which guarantees quantitative coherence". For example, when we consider a claim about the validities of the outcome of the next random experiment, while the frequentist conclude that the claim is "probably valid" or "probably not valid", the Bayesian may provide to what extent the "validity" is conditional on un-neglectable priori knowledge and data information.

While the frequentist approach is criticized for associating the probabilities to the data "too tight" based on the limiting relative frequency, the Bayesian associates the probabilities to the data and prior belief in a consistent manner simply based on Bayes' theorem and a few probability rules. If we have a belief in the model M, the initial information is I_0, and denote the new data information as y, the degree of our original belief, 'prior' $p(M \mid I_0)$, could be adapted to the 'posterior' $p(M \mid y, I_0)$ via the Bayes' theorem, which is expressed as the following:

$$p(M \mid I_0, y) \propto p(M \mid I_0)\, p(y \mid M)$$

where $p(y \mid M)$ is the probability density function of the new data conditional on model M. Because of the nature of the Bayesian approach, the posterior believes are adaptive and are always revised, which makes the inference logically consistent. Besides its adaptive characteristic, the Bayesian approach is operational and unified, unlike some non-Bayesian approaches that require special techniques and principles for different problems. Zellner (1971, pp.11) makes the point that "analyzing time series, regressions, simultaneous equation models in Bayesian framework all share the same approach and principles".

After clarifying some Bayesian fundamentals, it is important to compare the different ways that the frequentists and Bayesians solve problems such as estimating and hypotheses testing.

With respect to estimation, the frequentist focuses on the confidence interval estimation, while the Bayesians focus on the credible interval estimation. The frequentist believes that some fixed and unknown parameters exist objectively. For instance, we

assume the life time of batteries generated from a machine are identical and independent, following a normal distribution with a mean μ_0 and variance σ_0^2. Given the fact that we would not know the exact values of μ_0 and σ_0^2, we can estimate $\widehat{\mu}$ and $\widehat{\sigma}^2$ by randomly selecting a sample of batteries with a finite size. A $1-\alpha$ confidence interval of $\widehat{\mu}$ can be evaluated and expressed in a form of $\mu_1 < \mu < \mu_2$. Therefore, this estimated interval is a realization of a random interval from a distribution of confidence intervals. If we repeat the random sampling from the same population and calculate all the other confidence intervals in the same manner, the true mean μ_0 will lie in these confidence intervals a proportion $1-\alpha$ of the time. However, the statement that the true mean μ_0 has a $1-\alpha$ probability of being in an estimated confidence interval $\mu_1 < \mu < \mu_2$ is not appropriate, since the population mean μ_0 is either in this estimated confidence interval or not. In other words, the probability that a particular confidence interval contains the true mean μ_0 is either 1 or 0. Accordingly, we can only say the estimated interval "probably" covers the population mean μ_0 but we cannot tell what the probability is.

The Bayesian also believes that a true fixed value of the parameters exists. Because we do not know the true value of the parameters, μ_0 and σ_0^2 are viewed as random variables. A non-informative prior density, denoted as $p(\mu_0, \sigma_0^2)$, can be assigned to μ_0 and σ_0^2 as our state of ignorance before the data is obtained, where the non-informative indicates little "priori" knowledge. Given the data y, and a probability model $p(y \mid \mu_0, \sigma_0^2)$, we can calculate a $1-\alpha$ credible interval, which is sometimes referred as the highest posterior density interval (HPDI) according to the posterior density $p(\mu_0, \sigma_0^2 \mid y)$. Conditional on the "priori" statement and the data information, we can make a statement that this credible interval contains a proportion $1-\alpha$ of the total posterior probability. Therefore, when the non-informative prior is used, though the Bayesian credible interval and the frequentist confidence interval provide the same numerical answers, they have very different interpretations.

Similarly, for point estimations, when a non-informative prior is applied in the Bayesian framework, the frequentist provides identical numerical point estimates for most of the time. When the information is updated, the frequentist approach is often criticized for being incapable of "taking advantage of the additional knowledge" (Berger 1985, pp.139). On the contrary, the Bayesian approach "does reflect the benefits of the added knowledge"(Berger 1985, pp.139)[1]. For more details concerning the estimation issues in the frequentist and Bayesian approaches, readers may refer to

[1]Berger (1985, pp138-139) uses Example 7 to illustrate the differences of point estimation in frequentist and Bayesian.

Berger (1985, pp.140-145), Box and Tiao (1992, pp.84-86), Bernardo and Smith (1994, pp.259) and Zellner (1971, pp.24-28).

Regarding the hypotheses testing, we focus on the testing of a point null hypothesis. Hypothesis testing in the frequentist approach substantially differs from that in the Bayesian. In the frequentist method, hypothesis testings are involved with error probabilities, normally referred as Type 1 error and Type 2 error. Consider a situation, in which we want to test the unknown population mean μ equals to 0, presumably the population σ^2 is known as 1 and we choose a significance level at $\alpha = 5\%$. The following hypotheses are to be tested:

$$\begin{aligned} H_0 &: \quad \mu = 0 \\ H_1 &: \quad \mu \neq 0 \end{aligned}$$

Following the frequentist testing procedure, we can randomly collect a sample and calculate the P-value. The P-values are interpreted in a way that a small P-value indicates an evidence against the null hypothesis. The null hypothesis is rejected if P-value is smaller than a specified Type 1 error rate, 5% in this case. The frequentist will report that the hypothesis of $\mu = 0$ is rejected, that H_0 hypothesis is "probably" not true. However, we do not know how strong the evidence is against the H_0 hypothesis. Moreover, the classical approach encounters difficulties in testing non-nested hypotheses, as normally H_0 is required to be nested within H_1. Last but not least, frequentists are criticized in terms of philosophical reasoning. Given the nature of P-values, unobserved extreme numbers play important roles upon the decision making even if the these numbers did not occur[2]. Carlin and Louis (1996, pp.8) state ".. this violates the Likelihood Principle...frequentist test results actually depend not only on what is observed, but also on how the experiment was stopped".

In the Bayesian framework, the terminology to deal with hypotheses is "comparing" rather than "testing". Therefore, decision makers can have an option to choose either the H_0 or the H_1 based on the posterior probabilities $p(H_0 \mid y)$ and $p(H_1 \mid y)$, where the posteriors $p(H_0 \mid y)$ and $p(H_1 \mid y)$ are merely the updated prior belief of $p(H_0)$ and $p(H_1)$ conditional on the data information y. Since the posterior probabilities can be calculated numerically and represent degrees of belief in different propositions, the decision maker can then make explicit conclusions of not only which proposition is preferred but also to what extent the proposition is preferred in a quantitative

[2]Please see Carlin and Louis (1996, pp8) Example 1.2, based on two differently designed experiments, the P-value differs although the two experiment have identical likelihood functions.

manner. Furthermore, Bayesians do not suffer from problems when comparing non-nested models. Based on the quantified degree of belief, a group of propositions can be ranked and compared according to the posterior probabilities via Bayes Factors. This will be discussed in more details in Section 4 with respect to model selection in the Bayesian framework.

Talking about sharp point null hypothesis testing, the frequentist and Bayesian could provide entirely different answers. This is illustrated by Zellner (1971, pp.303-305). When the H_0 is rejected by the classical approach according to the conventional statistics at $P-value = 0.05$, the posterior probability $p(H_0 \mid y)$ could be substantially large that no evidence is provided against H_0 under certain specified priors. In a more extreme case, if the sample size becomes very large, the P-value will be so small that H_0 is rejected classically. However, the posterior probability $p(H_0 \mid y)$ could be close to 1 supporting the H_0 hypothesis. This surprising result is referred to "Lindley's paradox" or "Jeffrey's paradox". Details about hypotheses testing in the frequentist and Bayesian approaches can be found in to Berger (1985, pp.145-157), Zellner (1971, pp.291-306).

2.2 Prior Elicitations

This section reviews the prior elicitation methods and some conflicting views and arguments in the literature. In this thesis, all elicited priors are proper and most of them are conjugate priors. Thereby, it is necessary to explain briefly why and how the priors are elicited. Detailed prior elicitations for different entertained models shall be discussed in future Chapters respectively.

The nature of Bayesian inference relies on the prior belief $p(M \mid I_0)$ about various possible hypothesis and the un-neglectable data information y. A prior represents a "priori" belief that exists before the data is available. The priors play a central role in Bayesian analysis. Because the approach to choosing a prior distribution may differ among various researchers and different reasonable priors will result in different answers, Bayesian analysis is often criticized as being "subjective".

Bayesians respond to the criticism mainly in the following ways: (1) Select non-informative priors that are the least "subjective" so that inference in the posterior will solely depend on the current data information. (2) Carry out the prior sensitivity analysis, or "robustness analysis" by employing different priors. If all reasonable priors yield similar results, problems such as selecting what prior will receive less concern.

Berger (1985, pp.77-104) provides a variety of approaches to select both the non-

informative and informative priors. Non-informative priors can be applied when no priori information is available. Zellner (1971, pp.41-53) proposes choosing a diffuse prior and provides many examples of the use of non-informative priors. There are some criticism about the use of non-informative priors when they become improper. The improper priors are with infinite variances so that the probability density does not integrate to 1. First, it is argued that the Likelihood principles maybe violated by the use of improper priors (Geisser, 1984). Second, the "marginal posterior distributions are found to have an un-Bayesian property" (Dawid, Stone and Zidek (1973)) when improper priors are employed. This problem is known as the "marginalization paradox". Third, a non-integrability of the posterior may arise with an improper prior (see Wago, 2003), because we cannot always obtain a proper posterior when a non-informative prior is applied. In this thesis, to avoid non-integrability problems in the posterior, all priors applied are proper.

For computation simplicity, we would like to choose some prior distributions $p(\theta)$ that is conjugate to the likelihood $p(y \mid \theta)$. Under these priors, the posterior distributions $p(\theta \mid y)$ share the same distribution as the priors. Also, these priors have the same functional form as the likelihood function. These priors are referred to as natural conjugate priors. When the entertained model is highly dimensional, we can assign independent conjugate priors to each parameter. Accordingly, corresponding conjugate forms of each conditional posterior can be obtained. Even in high dimensional settings, the conjugate priors can produce unidimensional conditional posteriors in a closed form (see Carlin and Louis (2000, pp.27)). This feature of conjugate priors is very important when the Markov chain Monte Carlo integration techniques are heavily employed to construct conditional posterior distributions from the joint posterior distributions. Since the natural conjugate prior share the same functional form as the likelihood function, the interpretations of eliciting natural conjugate priors are reasonable. Koop (2003, pp.18) indicates that "the natural conjugate priors can be interpreted as arising from a fictitious data set from the same process that generate the actual data". Exponential families have conjugate priors, such as the normal distribution, Gamma distribution and the Chi-square distribution (see Morris, 1983). In this thesis, most of the priors elicited are conjugate priors.

Since we prefer to elicit proper priors that could adequately reflect available prior information, the assessment of prior density functions as to what values should be chosen in the corresponding informative conjugate prior distributions becomes crucial. Berger (1985, pp94-104) introduces the marginal distribution method to determine the prior distributions. Therefore, actual values can be applied to determine the priors.

We adopt the idea that priors should provide good predictions of what we think the data would look like. Thus, in this thesis, we carried out experiments and simulated data from the priors based on the parametric model to help eliciting the priors, based on the idea that fictitious data generated from the prior could mimic the actual data to a certain extent.

2.3 Bayesian Estimation

Section 1 has compared the estimation procedures in the frequentist and the Bayesian framework. This Section reviews the Bayesian estimations by focusing on a commonly applied numerical approximation method: Markov chain Monte Carlo (MCMC). First, we introduce the motivation of applying MCMC. Second, we present four conventional MCMC sampling algorithms, which are later on applied in Chapters 3, 4 and 5. Third, the MCMC diagnostic methods applied in this thesis are reviewed.

2.3.1 The Motivation of Applying Markov chain Monte Carlo

In the Bayesian framework, estimations for parametric models involves evaluations of marginal densities of each parameter from the joint posteriors density. Therefore, evaluation of integrals from the posterior conditionals provides the fundamentals for Bayesian inference. In low-dimension parametric models, the integrals can be obtained analytically. However, in high-dimensional models, complex integrals are hard to achieve in closed form, and thus must be evaluated numerically. Thanks to modern computing power, sampling-based methods have been dramatically developed in recent years to solve numerical integration problems in empirical Bayes analysis. The MCMC method for numerical approximations is commonly applied, where a posterior can be simulated from a certain Markov chain. Gelfand and Smith (1990) and Tanner and Wong (1987) point out that, in the context of empirical Bayes analysis, estimations via simulation has been virtually revolutionized by the Markov chain Monte Carlo method. Chib et.al (2002) offers the following illustration for the MCMC

> *..., the posterior distribution is sampled by simulation methods and the draws generated from the simulation are used to summarize the posterior distribution. The simulation is conducted by devising, and simulating, the transition density of an irreducible, aperiodic Markov Chain whose invariant distribution is the target posterior distribution.*

We are motivated to apply MCMC because of its obvious advantages, which are demonstrated in Geweke (1989), Carlin and Chib (1995): simulations via MCMC techniques could not only shed light on the distribution properties for any feature of interest, but also be diagnosed. We will elaborate the above statements at the end of this paragraph. Also, in Jun S. Liu (2001, pp.245) "one advantage of MCMC sampler is that the sampler can produce desirable random samples from a target distribution by making a series of local changes to an arbitrary initial state".

First, we introduce the advantage of MCMC applications in highly parameterized models. Compared to classical estimation, a distinct advantage of Bayesian estimation via the MCMC method is that simulation of posterior distributions is generally applicable to any model with complexity (Carlin and Chib 1995, Geweke 1989). In the classical framework, estimations can be carried out via maximum likelihood (see Sowell for example with an application to ARIMA, 1992), approximate maximum likelihood (Fox and Taqqu 1986, Li and McLeod 1986) and two-step procedures (Geweke and Porter-Hudak 1983). However, the frequentist always encounters problems if the parameter space in the entertained model is highly dimensional, especially when the estimations of latent variables are involved. Moreover, the maximum likelihood method is not always easy to conduct and estimations from the approximate maximum likelihood sometimes result in inconsistency. In the Bayesian framework, as long as the posterior moments exist, estimations can be achieved with the applications of MCMC, regardless of how highly parameterized the model is. Given enough replications, the numerical approximation of posterior moments converges almost surely to the true value, and the numerical accuracy of this numerical approximation can be diagnosed (see Carlin and Chib 1995, Geweke 1989). The MCMC convergence and accuracy diagnostics will be introduced in Section 3.3.

Second, simulations via MCMC techniques can also shed light on the posterior distribution properties. With the sampling outputs from a certain Markov chain, we can visualize the distribution of the parameter of interest by plotting the histograms of the sampled draws. Posterior moments and quantiles, mean that it is straight forward to make inference from the simulated distribution.

2.3.2 Sampling Algorithms

Substantive Monte Carlo strategies have been developed and applied in scientific computing, please refer to Jun S. Liu (2001) for detailed reviews. In this section, we provide brief reviews of some commonly applied MCMC algorithms, which are applied in this piece of research.

1. Gibbs Sampling Algorithm

The Gibbs sampler (Geman and Geman, 1984) is one of the most straightforward MCMC schemes to simulate the posterior distributions, where sequential draws are randomly drawn from the posterior conditionals. In this thesis, the Gibbs sampling algorithm will be applied extensively in Chapters 3 and 4.

In Liu (2001, pp.129-150) and Carlin and Louis (2000, pp.141-142), the Gibbs sampling algorithm is outlined as follows:

In the Bayesian context, we always encounter the joint posterior distributions $p(x_1, \cdots, x_k \mid y)$ with k variables, where the full or complete conditional distributions $p(x_i \mid x_{j \neq i}, y)$ with $i = 1, \cdots, k$, are available for sampling. However, our interest is in the marginal posteriors for $p(x_i \mid y)$,with $i = 1, \cdots, k$. Given an arbitrary set of starting values $\left\{x_1^{(0)}, \cdots, x_k^{(0)}\right\}$, the Gibbs sampling algorithm proceeds as follows

Get a random draw $x_1^{(1)} \sim p\left(x_1 \mid x_2^{(0)}, \cdots, x_k^{(0)}, y\right)$,

Get a random draw $x_2^{(1)} \sim p\left(x_2 \mid x_1^{(1)}, x_3^{(0)}, \cdots, x_k^{(0)}, y\right)$,

$$\vdots$$

Get a random draw$x_k^{(1)} \sim p\left(x_k \mid x_1^{(1)}, x_2^{(1)} \cdots, x_{k-1}^{(1)}, y\right)$

The first iteration of the scheme is completed when we have vector $\mathbf{x}^{(1)}$= $\left(x_1^{(1)}, \cdots, x_k^{(1)}, y\right)'$. After S iterations, we can have S vectors $\mathbf{x}^{(j)} = \left(x_1^{(j)}, \cdots, x_k^{(j)}, y\right)'$, where $j = 1, \cdots, S$. If we discard the initial S_0 vectors, the remained chain of $\left(x_i^{(S_0+1)}, \cdots, x_i^{(S)}\right)'$will be considered as valid random draws from $p(x_i \mid y)$, where $i = 1, \cdots, k$.

2. Metropolis-Hasting Algorithm

Very similar to the Gibbs Sampling Algorithm, the Metropolis-Hasting (M-H) algorithm is one of the most widely applied MCMC methods. The simplest form of the M-H algorithm is called the Metropolis algorithm, which was introduce by Metropolis et al. (1953). The Metropolis algorithm is generalized by Hastings (1970) and the algorithm is called the Metropolis-Hasting algorithm thereafter. We briefly summarize the M-H algorithm here (see Liu 2001, pp.115, Carlin and Louis, 2000, pp.152 and Koop 2003, pp.92 for detailed reviews).

The main feature of the M-H algorithm is a selection of a candidate generating density. The MCMC efficiency using M-H algorithm depends heavily on the choice of the candidate density.

Suppose the joint posterior density is denoted as $p(\mathbf{x} \mid y)$, where $\mathbf{x}$ =$x_1, \cdots, x_k$. We can select a candidate density denoted as $q(\cdot)$, from where candidate draws $\mathbf{x}^*$are

taken. Given an arbitrary set of starting values $\mathbf{x}^0 = \left(x_1^{(0)}, \cdots, x_k^{(0)}\right)'$, the M-H sampling algorithm is conducted as follows:

1. Take a candidate draw, $\mathbf{x}^*$ from the candidate generating density $q\left(\mathbf{x}^{(s-1)}; \mathbf{x}\right)$, which indicates the density $q\left(\mathbf{x}^{(s-1)}; \mathbf{x}\right)$ depending on the previous state in the Markov chain $\mathbf{x}^{(s-1)}$.

2. Calculate an acceptance probability,

$$\alpha\left(\mathbf{x}^{(s-1)}, \mathbf{x}^*\right) = \min\left[\frac{p\left(\mathbf{x} = \mathbf{x}^* \mid y\right) q\left(\mathbf{x}^*; \mathbf{x} = \mathbf{x}^{(s-1)}\right)}{p\left(\mathbf{x} = \mathbf{x}^{(s-1)} \mid y\right) q\left(\mathbf{x}^{(s-1)}; \mathbf{x} = \mathbf{x}^*\right)}, 1\right],$$

   where $p\left(\mathbf{x} = \mathbf{x}^* \mid y\right)$ is the posterior density evaluated at $\mathbf{x}^*$, while $q\left(\mathbf{x}^*; \mathbf{x} = \mathbf{x}^{(s-1)}\right)$ is the $q(\cdot)$ density evaluated at $\mathbf{x} = \mathbf{x}^{(s-1)}$.

3. Set $\mathbf{x}^{(s)} = \mathbf{x}^*$ with probability $\alpha\left(\mathbf{x}^{(s-1)}, \mathbf{x}^*\right)$ and set $\mathbf{x}^{(s)} = \mathbf{x}^{(s-1)}$with probability $1 - \alpha\left(\mathbf{x}^{(s-1)}, \mathbf{x}^*\right)$.

4. Do steps 1, 2 and 3 S times and discard the initial S_0 draws.

Other evolutive M-H algorithms have been developed based on the M-H sampling algorithm, such as the Independent Chain M-H sampling. The Independent Chain M-H sampling does not require the candidate generating density depending on $\mathbf{x}^{(s-1)}$. When a convenient approximation to the posterior exists, the convenient approximation can be used as a candidate generating density. Then, the acceptance probability simplifies to

$$\alpha\left(\mathbf{x}^{(s-1)}, \mathbf{x}^*\right) = \min\left[\frac{p\left(\mathbf{x} = \mathbf{x}^* \mid y\right) q^*\left(\mathbf{x} = \mathbf{x}^{(s-1)}\right)}{p\left(\mathbf{x} = \mathbf{x}^{(s-1)} \mid y\right) q^*\left(\mathbf{x} = \mathbf{x}^*\right)}, 1\right]$$

3. Griddy-Gibbs Sampling Algorithm

The Griddy-Gibbs sampling algorithm can be applied when the full conditional posterior density is not available. Griddy-Gibbs Sampling is extensively discussed in Ritter and Tanner (1992). Bauwens and Lubrano (1998) show that the Griddy-Gibbs sampler works well in practice with GARCH models. Here, we briefly introduce the Griddy-Gibbs sampling algorithm and its areas of application.

The idea of Griddy-Gibbs sampling is simple: Suppose we have a non-standard posterior that we wish to obtain random draws from, the density can be approximated on a grid of points. Then, we can apply the Inversed cdf sampling method to obtain

random draws from the approximated posterior density. The Griddy-Gibbs sampling algorithm is conducted as follows (please see details in Ritter and Tanner (1992)):

1. Set the number of grid points as n and evaluate posterior densities $p\left(X_i|X_j,y\right)$, where $i \neq j$, at n number of discrete points $X_i = x_1, x_2, \cdots, x_n$. Then, we have the posterior densities $w_1, w_2, \cdots, w_n$ evaluated at $x_1, x_2, \cdots, x_n$.

2. Use $w_1, w_2, \cdots, w_n$ to obtain an approximation to the inverse cdf of $p\left(X_i|X_j,y\right)$, where $i \neq j$.

3. Get a random draw from a uniform distribution $u \sim f_U\left(0,1\right)$.

4. Find the point x_i, at which the approximated cdf corresponding to $p\left(X_i|X_j, i \neq j\right)$ equals to u. Then, x_i is a valid random draw from $p\left(X_i|X_j,y\right)$.

There are 5 important remarks to be made regarding to the Griddy-Gibbs sampling algorithm, which are summarized as follows:

(1) The function $p\left(X_i|X_j, i \neq j\right)$ needs be known only up to a proportionality constant.

(2) The grid $x_1, x_2, \cdots, x_n$ need not to be uniformly spaced.

(3) The number of points in the grid need not to be constant over the iterations of the Gibbs sampler.

(4) Over an unbounded interval it is important to compare w_1 and w_n to $M = \max(w_1, \cdots, w_n)$ If $w_1 > f \cdot M$, where $0 < f < 1$, the grid must be augmented with points to the left of x_1; if $w_n > f \cdot M$, the grid must be augmented with points to the right of x_n.

(5) Simple approximations to the inverse cdf are (a) Piecewise constant corresponding to a discrete distribution for $p\left(x_i\right) = \frac{w_i}{\sum_{j=1}^{n} w_j}$ (b) Piecewise linear corresponding to a piecewise uniform distribution on the interval $[a_i, a_{i+1}]$, $i = 1, \cdots, n-1$ with $x_i \in [a_i, a_{i+1}]$ and density $f_i = \frac{w_i}{\sum_{j=1}^{n} \omega_j}$, where $\omega_j = w_i \cdot (a_{i+1} - a_i)$. Typically, x_i is the mid-point in the interval$[a_i, a_{i+1}]$.

4. Other Sampling Algorithms

The sampling algorithms introduced above can be easily applied in practice. For other sampling algorithms and their applications, please refer to Liu (2001), which provides detailed and extensive reviews of Monte Carlo strategies.

2.3.3 MCMC Diagnostics

Since the MCMC algorithms are in fact plans for moving a point around its sample space, a proper MCMC algorithm should satisfy the following two criteria: (1) We have to make sure the point's movement leaves the target distribution invariant. (2) Further to this, we have to provide sufficient samples so that the target distribution could be represented and simulated, where "sufficient" means that at a time T the MCMC algorithm has converged. In Carlin and Louis (2000, pp.172), if an MCMC algorithm converges at time T, "no pre-convergence, or burnt-in, samples should be retained for subsequent inference". According to Carlin and Louis, in practice, "safely" is taken as "no pre-convergence samples should be retained for subsequent inference", where the pre-convergence samples is also often referred as burnt-in samples. In other words, the stopping rule for a MCMC algorithm is that the convergence of MCMC must be achieved after the burn-in. Carlin and Louis (2000, pp.172) demonstrate why a MCMC algorithm may be expected to converge using a simple bivariate two-compartment model.

With respect to MCMC diagnostics, one of the most frequently cited papers is Geweke (1992), titled "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments". Geweke's idea lies in the assumption that the chain can be treated as a time series. Concerning various approaches to diagnosing the convergence in the MCMC, Cowles and Carlin (1996) provides a comprehensive review of the Markov chain Monte Carlo convergence diagnostics in the Journal of the American Statistical Association.

Following Carlin and Louis (2000, chapter 5.4.6), in this section, we provide a brief review and references of the (a) common causes of convergence failure, (b) diagnostic tools used to make stopping decisions for MCMC algorithms. For understandings of Monte Carlo Strategies in depth, Liu's book (2001) "Monte Carlo Strategies in Scientific Computing" offers comprehensive reviews in this regard.

(a) In general, the convergence failures are mostly due to the high autocorrelations in the realized sample chains. The high autocorrelations and slow convergence normally occur when the model is so large that the model is overparameterized. In Carlin and Louis (2000, pp.175 line 10), "Overparameterizations typically lead to high posterior correlations among the parameters and, therefore, high crosscorrelations will retard the movement of the Gibbs sampler through the parameter space".

LeSage (1998, pp.161) stress that " a high degree of autocorrelation indicates that we may need to carry out more draws to achieve a sample of sufficient size to draw accurate posterior estimates".

(b) As Carlin and Louis (2000, PP.177 line 11) pointed out, theoretical convergence bounds and perfect sampling strategies[3] are difficult in practical use, "almost all of the applied work involving MCMC methods has relied on the applications of diagnostic tools to output produced by the algorithm". References of a few of other methods can also be found in Carlin and Louis (2000, pp.178).

However, we have to emphasize that no diagnostic can "prove" convergence of a MCMC algorithm. Because whatever the diagnostic is, it only use a finite realization from the chain. Cowles and Carlin (1996) summarize that all diagnostics are fundamentally unsound. We use the diagnostic tools to analyze the convergence in the MCMC, same as other statisticians, simply because that "a weak diagnostic is better than no diagnostic at all" (Carlin and Louis, 2000, pp.178 line 12).

There are various methods to see whether the estimated results from the MCMC are reliable. Different MCMC diagnostics are applicable in regards to their diagnostic goals and dimensionality of the posteriors. Carlin and Louis (2000, pp.179) reviewed several diagnostics with respect to different characteristics. In the end, they summarized the diagnostic strategy, which was recommended by Cowles and Carlin (1996).

Following the recommendation in Carlin and Louis (2000, pp.182) in general, the MCMC diagnostics are carried out with the following strategy:

1. Run 3-5 parallel chains with different starting points drawn from a distribution believed to be overdispersed with respect to the stationary distribution (say, covering ± 3 prior standard deviations from the prior mean). The first $10\% - 20\%$ of the chain is discarded as the burnt-in, whereas the rest of the chain is retained for inference (For discussions in regard to overdispersed initial conditions and burn-in, please refer to Gelman and Rubin (1992)).

2. Inspect the chains by plotting the sampled values from the MCMC algorithm.

3. Provide the correlograms of the sample draws in a sense that correlation rates are a useful tool for MCMC efficiency diagnostics. In Liu (2001, pp.130), as the sequential draws from the MCMC are correlated, the convergence speed depends on the correlation rates between the iterations. In practice, putting variables that are highly correlated into blocks can help to improve efficiency, e.g. MCMC sampling scheme for Stochastic Volatility models developed by Chib et al. (2002).

[3] In Carlin and Louis (2000, pp.176), "perfect sampling" refers to MCMC simulation methods, which can guaranee that a sample drawn at a given time will be exactly distributed accoridng to the chain's stationary distribution.

4. Calculate Numerical Standard Errors (NSE), Relative Numerical Efficiency (RNE) and Convergence Diagnostic (CD) values using the retained samples. **NSE** is a measure of the precision of the Monte Carlo estimate. In Geweke (1992) section 3.1, given the idea that the chain can be treated as a time series $\{G(j)\}$ with power spectrum $S_G(\varpi)$, the corresponding numerical standard error (NSE) of the estimate is: $NSE = \left[R^{-1}\widehat{S_G}(0)\right]^{1/2}$ where $R^{-1}S_G(0)$ is the asymptotic variance of $\{G(j)\}$, $S_G(0)$ is the spectrum of $\{G(j)\}$at 0 frequency, estimated as $\widehat{S_G}(0)$ and R is the Monte Carlo sample size. According to LeSage (1998, pp.164), "**RNE** estimates provide an indication of the number of draws that would be required to produce the same numerical accuracy if the draws represented had been made from an iid sample drawn directly from the posterior distribution." Geweke (1992) promotes the use of the RNE as a measure for the quality of a correlated sample, where RNE is the ratio between the empirical variance of the sample and a correlation-consistent variance estimator. According to Geweke (1992), $RNE = \widehat{var}[g(\theta)]/\widehat{S_G}(0)$ .The RNE does not depend on the number of iterations taken from the sampler. "Positive serial correlation of $\{G(j)\}$ renders the Gibbs sampling estimator less efficient, and negative serial correlation in $\{G(j)\}$ renders it more efficient" (Geweke, 1992). **CD** is assessing the convergence of the chain. The CD value is also proposed by Geweke (1992), the MCMC is converged as long as the absolute CD value is smaller than 1.96. In Koop (2003, pp.66), the "CD value can be calculated and compared to critical values from a standard Normal statistical table". Also, in Geweke (1992), equation 3.1. The formula of calculating CD values are given on pp.33. (For details of the usage of NSE, RNE and CD, please see Geweke (1992), Koop (2003, pp.64-68) and Appendix A).

In this thesis, MCMC algorithms are diagnosed following the above steps.

2.4 Model Selections

Model selection should best be viewed as a decision problem (Draper 1999 and Hoeting et al. 1999). In the Bayesian context, we review the model assessment and selection procedure in this section. Regarding questions such as how the selection of priors affect the model selection decision, how to compare entertained models based on the data information and how to make the ultimate decision in the model selection. Some computational methods involved for model assessment are reviewed, and are intensively applied in future chapters.

First, model assessment in the Bayesian framework is achieved by the Bayes Factors B_{ij} (choice between model M_i and model M_j), which is also known as the marginal likelihood ratio

$$B_{ij} = \frac{p\left(y \mid M_i\right)}{p\left(y \mid M_j\right)}$$

where $p\left(y \mid M_i\right)$ and $p\left(y \mid M_j\right)$ are the marginal likelihoods in the light of available data based on model M_i and M_j respectively. Any models of interest can be compared as long as their marginal likelihoods are achievable. Therefore, using the marginal likelihood for model selection is conceptually the simplest. Further, rather than presenting pairwise comparisons, the fit of the entertained models in any form can be inferenced via the marginal likelihood, while these inference or tests in the classical framework are always difficult to attain.

In practice, the ultimate decision of model selection among a group of entertained models will depend on the model's marginal likelihood based on the data.

Bayes Factors for model comparison have many advantages. Interested readers are directed to Kass and Raffrey (1995) who provide comprehensive reviews of Bayes factors. In regard to the nonlinearity issue in time series modelling, Koop and Potter (1999) state that, "the Bayesian is forced to specify the types of nonlinearity that he/she expects to occur and only if the nonlinearity occurs as expected will the Bayes Factors favour the nonlinear model... This is a positive development since it forces the econometricians to think carefully about the underlying economic theory"[4].

The connection between the Bayes factor B_{ij} and the posterior odds ratio $\frac{p(M_i|y)}{p(M_j|y)}$ are expressed as the following

$$\frac{p\left(M_i|y\right)}{p\left(M_j|y\right)} = \frac{p\left(y \mid M_i\right)p\left(M_i\right)}{p\left(y \mid M_j\right)p\left(M_j\right)} = B_{ij}\frac{p\left(M_i\right)}{p\left(M_j\right)} \tag{2.1}$$

where $\frac{p(M_i)}{p(M_j)}$ is known as the prior odds ratio. If all models are treated as "priori" equally likely, the prior odds ratio equals one. Consequently, the posterior odds ratio equals the Bayes factor. In the cases that the prior odds ratio does not equal unity, such that we have different preferences over different models in the prior belief, a choice between model i and j should be based on calculations of the Bayes factors.

The strength of evidence in favours of model M_i versus M_j is evaluated according to the Bayes factor scale in Table (2.1) under Jeffreys's classifications:

[4]These issues are discussed on a more general level in Poirier (1995), Chapter 10.

Table 2.1: Jeffreys Interpretation of Bayes Factors

$B_{ij} > 1$	Support for M_i
$10^{-\frac{1}{2}} < B_{ij} < 1$	Very slight evidence against M_i
$10^{-1} < B_{ij} < 10^{-\frac{1}{2}}$	Slight evidence against M_i
$10^{-2} < B_{ij} < 10^{-1}$	Strong to very strong evidence against M_i
$B_{ij} < 10^{-2}$	Decisive evidence against M_i

Since the Bayes factor is the fundamental tool for model selection in the Bayesian context, it is intuitive to calculate the marginal likelihood of all entertained models and calculate the model probabilities. Using the compound probability rule, the marginal likelihood can be calculated by numerical integration

$$p\left(y \mid M_i\right) = \int_{\Theta} p\left(y \mid M_i, \Theta\right) p\left(\Theta \mid M_i\right) d\Theta \tag{2.2}$$

where Θ denotes the parameters in M_i. However, when M_i is highly parameterized, the calculation from (2.2) will rarely be available in closed form. However, modern computational methods (i.e. the MCMC method) facilitate the development of applied Bayesian in the context of nonlinear modelling. For many high dimensional nonlinear models, estimations always involve MCMC in practice, which require some form of sampling to evaluate the posterior distributions of interest. This means that common model checks will be available at little extra cost when the posterior simulator is available.

The calculations of Bayes factors via marginal density estimations have been reviewed in Kass and Raftery (1995), Carlin and Louis (2000, pp.204-210), Han and Carlin (2001), which include all of the computational methods described below.

1. Monte Carlo Integration with Observations from the Prior

It is often referred to a direct method, when we generate observations directly from the prior to estimate the entertained model's marginal likelihood. Recall the numerical expression of marginal likelihood in (2.2), We could generate $G : g = 1, \cdots, G$ observations of $\left\{\theta_j^{(g)}\right\}$ from the prior $p\left(\theta_j\right)$ and estimate the following

$$\widehat{p}\left(y \mid M_j\right) = \frac{1}{G}\sum_{g=1}^{G} p\left(y \mid \theta_j^{(g)}, M_j\right),$$

As remarked in Carlin and Louis (2000, pp.207), this direct Monte Carlo integration method is normally very unstable and inefficient, due to the fact that the conditional likelihood $p(y \mid \theta_j, M_j)$ can be very peaked compared to the prior $p(\theta_j)$, which leads to most of $p\left(y \mid \theta_j^{(g)}, M_j\right)$ being near 0 and a very inefficient estimator of $\widehat{p}(y \mid M_j)$. As shown in Fig (2-1), in a case that $5,353$ out of $6,000$ of $\theta_j^{(g)}$ from $p(\theta_j)$ fall in a low-likelihood region, which make most of $p\left(y \mid \theta_j^{(g)}, M_j\right)$ equals to 0. The estimate of $\widehat{p}(y \mid M_j)$ is dominated by 647 large values of the likelihood.

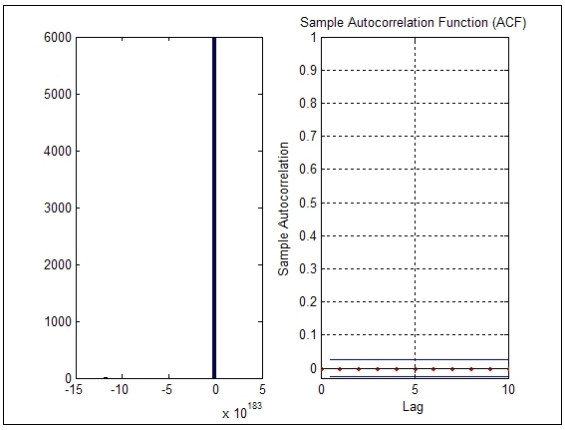


Figure 2-1: log Likelihood Using Samples from the Prior

The graph on the left hand side plots the histogram of $p\left(y \mid \theta_j^{(g)}, M_j\right)$ and the graph on the right hand side plots the correlograms of $\theta_j^{(g)}$. Bos (2002), provides a summary of various approaches to calculating marginal likelihood.

2. Harmonic Mean Method

A generalized Harmonic Mean (HM) method is proposed by Newton and Raftery (1994),

$$\widehat{p}(y \mid M_j) = \frac{\sum_{g=1}^{G} w^{(g)} \cdot p\left(y \mid \theta_j^{(g)}, M_j\right)}{\sum_{g=1}^{G} w^{(g)}}$$

where the Central Limit Theorem is satisfied. In this case, $\theta_j^{(g)}$ are sampled from an importance density $p^*(\theta_j)$, calculated as:

$$p^*(\theta_j) = \delta p(\theta_j) + (1-\delta) p(\theta_j \mid y)$$

where $p(\theta_j)$is the prior density and $p(\theta_j \mid y)$ is the posterior density of the entertained

model (M_j) respectively, and $w^{(g)} = p\left(\theta_j^{(g)}\right)/p^*\left(\theta_j^{(g)}\right)$. According to Newton and Raftery (1994), δ can be chosen as a small number.

3. Gelfand-Dey Method

Gelfand and Dey (1994) provide another approach to calculating the marginal likelihood. If the inverse of the marginal likelihood $p(y \mid M_j)$for model M_j only depends upon a parameter vector θ_j, this inversed marginal likelihood can be written as

$$\frac{1}{\widehat{p}(y \mid M_j)} = E\left[g(\theta_j) \mid y\right]$$

for a particular choice of $g(\theta_j)$,

$$g(\theta_j) = \frac{f(\theta_j)}{p(\theta_j \mid M_j)\, p(y \mid \theta_j, M_j)}$$

where $f(\theta_j)$ is any p.d.f with support contained in $\widehat{\Theta}$. For proof, please refer to Koop (2003, pp.104-106, section 5.7). To insure that the Gelfand-Dey method works well, $f(\theta_j)$ must be selected with caution.

To choose $f(\theta_j)$, we follow the recommendations in Geweke (1999), which has been found to work well in practice. This strategy involves letting $f(\theta_j)$ be a Normal density with the tails chopped off. The justification for this truncation is that the Gelfand-Dey method requires $\frac{f(\theta_j)}{p(\theta_j|M_j)p(y|\theta_j,M_j)}$to be finite for every possible value of θ_j. However, it is often difficult to verify whether $\frac{f(\theta_j)}{p(\theta_j|M_j)p(y|\theta_j,M_j)}$ is also finite in the tails of the Normal density. Therefore, $f(\theta_j)$ is set to zero in these (tails) potentially problematic regions.

Formally, let $\widehat{\theta}_j = E(\theta_j \mid y, M_j)$ and $\widehat{\sum} = var(\theta_j \mid y, M_j)$, which can be obtained from the posterior simulator. Further, for some probability, $p \in (0,1)$, let $\widehat{\Theta}$ denote the support of $f(\theta_j)$ which is defined by

$$\widehat{\Theta} = \left\{\theta_j : \left(\theta_j - \widehat{\theta}_j\right)' \widehat{\sum}^{-1} \left(\theta_j - \widehat{\theta}_j\right) \leq \chi^2_{1-p}(k)\right\}$$

where $\chi^2_{1-p}(k)$ is the $(1-p)th$ percentile of the Chi-squared distribution with k degrees of freedom and k is the number of elements in θ_j. Then Geweke (1999) recommends letting $f(\theta_j)$ be the multivariate Normal density truncated to the region $\widehat{\Theta}$,

$$f(\theta_j) = \frac{1}{p(2\pi)^{\frac{k}{2}}} \left|\widehat{\sum}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta_j - \widehat{\theta}_j\right)' \widehat{\sum}^{-1} \left(\theta_j - \widehat{\theta}_j\right)\right] 1\left(\theta_j \in \widehat{\Theta}\right)$$

where $1(\cdot)$ is the indicator function. In Koop (2003, pp.106), p can be selected as a small value, e.g 0.01, or a range of small values, since "there is almost no additional cost" (Geweke 1999) in doing so. To evaluate the accuracy of the estimated $\widehat{p}(y \mid M_j)$, NSE values can be calculated in a standard way (see Chapter 2, Section 3.3 for discussions in regard to NSE).

In general, the Gelfand-Dey method for estimating $\widehat{p}(y \mid M_j)$ works for any kind of models. However, there are some limitations of the Gelfand-Dey method: First, Carlin and Louis (2000, pp.208) point out that if θ_j is with a very high dimension, it will be difficult to choose an appropriate density $f(\theta_j)$, thus the Gelfand-Dey method may be incompetent to estimate $\widehat{p}(y \mid M_j)$. Second, Koop (2003, pp.106) stresses that the Gelfand-Dey method requires the full p.d.f of $p(\theta_j \mid M_j)$, $p(y \mid \theta_j, M_j)$ and a posterior simulator. In the cases that only kernels of the prior and/or likelihood is/are known, the Gelfand-Dey method cannot be applied. Finally, If there is a small number of $\theta_j \in \widehat{\Theta}$, greater simulation error could occur.

The follows discusses the methods using samples from the prior, Harmonic mean method, where the estimate of marginal likelihood is denoted as m (Bos, 2002):

> ...a simulation method may help. The method of m_{prior} is very unstable and inefficient. The method m_{IS} is not operational without a choice for the importance sampling density $\pi^*(\theta)$, approximating the prior density. Using the prior density as importance function leads to weights $w_i \equiv 1$ as in m_{prior}, but many drawings will fall in low-likelihood regions. Sampling from the posterior density gives more drawings in the correct region, but leads to an estimate of m which may not have a finite variance (Newton and Raftery 1994). Intermediate positions, like $\pi^*(\theta) = \delta\pi(\theta) + (1-\delta)\, p(\theta \mid y)$ can be chosen: This gives a consistent estimate with better convergence behaviour (see also Newton and Raftery 1994). A more recent solution for stabilizing the harmonic mean estimate, utilizing a technique of lowering the dimension of the problem, is given in Satagopan, Newton and Raftery (2000).
>
> Both m_{prior} and m_{HM} methods are special versions of m_{IS}.

The Gelfand-Dey estimator (see Gelfand and Dey, 1994 for details) is commonly known as the modified harmonic mean estimator in the literature.

4. Chib Method: Using Gibbs Sampler Output

Chib (1995) provided a method, which is therefore referred to as the Chib method. This method uses the Gibbs sampler with closed form full conditional distributions,

to calculate the marginal likelihood of the entertained model. In Chib (1995), the method begins with the marginal likelihood identity.

According to the Bayes' rule, the marginal density $p(y \mid M_j)$ can be written as

$$p(y \mid M_j) = \frac{p(y \mid \theta_j, M_j)p(\theta_j \mid M_j)}{p(\theta_j \mid y, M_j)}$$

where the numerator is the product of the sampling density and the prior, and the denominator is the posterior density of θ_j, with all the integrating constants included. As the above identity holds for any θ_j, say θ_j^*, the value of posterior density $p(\theta_j^* \mid y)$ can be estimated as $\widehat{p}(\theta_j^* \mid y)$ using Monte Carlo samples.

Therefore, the log marginal likelihood can be estimated as

$$\ln \widehat{p}(y) = \ln p(y \mid \theta_j^*) + \ln p(\theta_j^*) - \ln \widehat{p}(\theta_j^* \mid y)$$

at an arbitrary value of θ_j^*. While $\ln p(y \mid \theta_j^*)$ and $\ln p(\theta_j^*)$ are available straight forward from the Gibbs sampler outputs, as long as we know the exact form of the $p(\theta_j)$ and the $p(y \mid \theta_j)$, the estimate of $\ln \widehat{p}(\theta_j^* \mid y)$ involves a "reduced Gibbs run".

Take the same case in Carlin and Louis (2000, pp.209) as an example, suppose model M_j is a simple model with only two dimensions in θ_j, where $\theta_j = (\theta_{j,1}, \theta_{j,2})$, then the joint posterior at θ_j^* will be

$$p(\theta_{j,1}^*, \theta_{j,2}^* \mid y) = p(\theta_{j,2}^* \mid y, \theta_{j,1}^*) p\left(\theta_{j,1}^* \mid y\right)$$

while $p\left(\theta_{j,1}^* \mid y\right)$ can be estimated via the "Rao-Blackwellized" mixture estimate as follows:

$$\widehat{p}\left(\theta_{j,1}^* \mid y\right) = \frac{1}{G}\sum_{g=1}^{G} p\left(\theta_{j,1}^* \mid y, \theta_{j,2}^{(g)}\right) \tag{2.3}$$

since $\theta_{j,2}^{(g)} \sim p\left(\theta_{j,2} \mid y\right) : g = 1, \cdots, G$. More explicitly, we can use the output of $\theta_{j,2}^{(g)} : g = 1, \cdots, G$ from the Gibbs sampler to estimate $\widehat{p}\left(\theta_{j,1}^* \mid y\right)$ in (2.3) at $\theta_{j,1}^*$. Therefore, if the parameters in θ_j are divided into B blocks, we need to run the Gibbs sampler $B-1$ times consecutively to estimate $\ln \widehat{p}(\theta_j^* \mid y)$.

The Chib method can be applied virtually for any type of model of interest for marginal likelihood estimation. One disadvantage of the Chib method is the computational burden. Since this algorithm demands a huge amount of iterations, the execution time will be long if the dimension of θ_j is very high. To reduce sampling time and improve computational accuracy, it will be necessary to partition θ_j into

Table 1: Computational methods for marginal likelihoods

$m_{\text{Anal}} = \int \mathcal{L}(Y,\theta)\pi(\theta)\partial\theta$
The exact solution, available only in special cases.

m_{Num} Using brute-force numerical integration, m can be computed for low-dimensional problems.

$m_{\text{Prior}} = \mathsf{E}_{\pi(\theta)}\mathcal{L}(Y,\theta) \approx \frac{1}{n}\sum \mathcal{L}(Y,\theta^{(i)}); \theta^{(i)} \sim \pi(\theta)$
Simulating from the prior, m is the average likelihood. A very unstable and inefficient estimate.

$m_{\text{IS}} = \frac{1}{\sum w_i}\sum w_i \mathcal{L}(Y,\theta^{(i)}); \theta^{(i)} \sim \pi^*(\theta), w_i = \pi(\theta^{(i)})/\pi^*(\theta^{(i)})$
Sampling from an importance density with higher probability mass around the posterior mode improves on m_{Prior}.

$m_{\text{HM}} = \left(\frac{1}{N}\sum \frac{1}{\mathcal{L}(Y,\theta^{(i)})}\right)^{-1}; \theta^{(i)} \sim p(\theta|Y)$
Using the posterior density as importance sampling density in m_{IS} leads to using the harmonic mean of the likelihood values as estimator.

$m_{\text{App}} = \mathcal{L}(Y,\theta)\pi(\theta)/p(\theta|Y)$
In general, relating the height of the posterior kernel to the height of an approximating posterior density in a high density point θ can give an estimate of m.

$m_{\text{LP}} = (2\pi)^{k/2}|\tilde{\Sigma}|^{\frac{1}{2}}\mathcal{L}(Y,\tilde{\theta})\pi(\tilde{\theta})$
This is m_{App} using a quadratic expansion of the posterior kernel around the posterior mode $\tilde{\theta}$.

$m_{\text{Kern}} = \mathcal{L}(Y,\theta)\pi(\theta)/p_{\text{Kern}}(\theta|Y)$
This is m_{App} using a kernel smoother to approximate the posterior density at a high density point θ.

$m_{\text{Gibbs}} = \mathcal{L}(Y,\theta)\pi(\theta)/\left[p(\theta_1|Y)\prod p(\theta_i|\theta_1,..,\theta_{i-1},Y)\right]$
Chib (1995) proposed to use $k-1$ samples from separate Gibbs chains to approximate $p(\theta_i|\theta_1,..,\theta_{i-1},Y) \approx \sum p(\theta_i|\theta_1,..,\theta_{i-1},\theta_{i+1}^{(j)},..,\theta_k^{(j)},Y)/N$, where N samples are drawn for the missing elements $\theta_{i+1},..,\theta_k$ from the conditional densities.

Figure 2-2: Marginal Likelihood evaluations from Bos (2002)

blocks (see Chib et al. 2002).

5. Other Methods

Many scholars have been concerned with the computational issues in regard to marginal likelihood for model comparisons. Besides the four methods introduced above, Bos (2002) reviews other methods that are commonly applied in practice, such as the LaPlace method and Kernel smoother method (see attached Figure 2-2).

Also, the Chib method has been extended in Chib and Jeliazkov (2005) to estimate the marginal likelihood via Metropolis-Hastings output. In this thesis, since the Gelfand-Dey and the Chib method are intensively applied in Chapter 3 and Chapter 4, we do not provide lengthy reviews of other methods. For other methods of Bayes factor calculations, such as Savage-Dickey density ratio method please refer to (Verdinelli and Wasserman (1995), Koop (2003, pp.81-82)), sampling over the model and parameter space method, please refer to Carlin and Louis (2000, pp.211-219).

2.5 Bayesian Forecasting and Bayesian Model Averaging

In this section, we aim to provide a brief review of Bayesian Model Averaging (BMA), a distinctive approach in Bayesian context, when we inevitably encounter the forecasting uncertainties.

Forecasting, in a Bayesian context, is carried out based on the predictive densities $p(y^* \mid y)$, where the future y^* are mapped from the past y through conditional probabilities that all uncertainties are (or should be) represented. Zellner (1971, Chap 2), Bernado and Smith (1994, Chap 2), and Berger (1985, Chap 4, Chap 7, Chap8), Carlin and Louis (2000, Appendix B) offer very general ideas of Bayesian forecasting and Bayesian decision making theories. West and Harrison (1997) offers very systematic introductions of Bayesian modelling and Bayesian forecasting in regards to various dynamic models. Bayesian forecasting methods have great potential due to the fact that "they provide tractable solutions for problems that prove difficult when approached using non-Bayesian methods" (Geweke, 2001).

Combining forecasts has strong advocates in both the Bayesian and Classical framework. Bates and Granger (1969) is an influential paper in the combining forecasts literature, which argues that a combination of two models can give a better forecast than either individual model. Inference from a single model alone is often considered risky, since "...conditional on the chosen model would of necessity ignore the uncertainty in the model selection process itself and the process of model selection may be somewhat arbitrary in that a number of models may fit equally well and simultaneously provide plausible explanations of the data"(see Carlin and Louis 2000, pp.49). Madigan and Raftery (1994) support the idea of model averaging and show that averaging over all models can provide better predictive ability using the nonnegativity of the Kullback-Leibler divergence. Therefore, questions, such as how to combine the forecasting models, are naturally raised. Researchers propose different ways to combine forecasts (see Draper 1995, and Min and Zellner 1993). In regards to this area, Clemen (1989), and de Menezes, Bunn, and Taylor (2000) provides intensive reviews of historical research from the 60's to the 90's. The area of combining forecasts can be fractionalized into different divisions, such as combining event (Clements, 2006) and quantile forecast (Clements et al., 2006), combining density forecasts in Hall and Mitchell (2004, 2007) and Clements (2004).

In recent years, density forecasts have become very attractive in the sense that all uncertainties associated with the forecast can be presented. The Bayesian approach

became very appealing for the following reasons:

1. The whole predictive density could be represented with no extra cost when the posterior simulator is available.

2. Bayes factors can be used to construct model probabilities for any kind of entertained models.

3. Furthermore, BMA provides a good solution to combining forecast densities based on a simple idea: many models are plausible representations of the data and models should be weighted averaged according to their posterior model probabilities. Hoeting et al. (1999) provides an excellent historical perspective and summarization of the literature on BMA.

Hoeting et al. (1999) also provides insightful examples in a wide variety of applications, demonstrated that BMA provides much better predictive performance than using a single model. BMA appears to have worked well in practice with monetary policy (Milani 2004, Garratt et al. 2007), consumption and wealth effects (Potter et al., 2005), forecasting Swedish inflation (Eklund and Karlsson, 2005, Karlsson and Jacobson, 2004) and US exchange rates (Koop and Potter, 2003).

For a coherent systematic analysis accounting for model uncertainties, we summarize that the BMA can be carried out with the following procedure:

1. We start with specifying the model space, where K selected candidate models $(M_1, \cdots, M_k)$ are contained. Elicit priors for Φ, which is a vector of common parameters across all the models.

2. In Raftery et al. (2003), using a few simple probability rules, the forecast pdf for given data y, $p(y^* \mid y)$, is given by

$$p(y^* \mid y) = \sum_{k=1}^{K} p(y^* \mid y, M_k)\, p(M_k \mid y)$$

where $p(y^* \mid y, M_k)$ is the forecast pdf based on statistical model M_k and data information, and $p(M_k \mid y)$ is the posterior model probability of M_k. If we denote $\varpi_k = p(M_k \mid y)$, since the posterior model probabilities of all candidate models adds up to 1, we have

$$\sum_{k=1}^{K} \varpi_k = 1$$

where ϖ_k can be evaluated via Bayes factors by model comparison amongst all the entertained models.

Intuitively, posterior model probabilities can be obtained straight forwardly as the following:

$$\varpi_k = \frac{p\left(y \mid M_i\right)}{\sum_{k=1}^{K} p\left(y \mid M_k\right)}$$

where $p\left(y \mid M_k\right)$ is the marginal likelihood of M_k. However, this method only works when K is small. In this thesis, we focus on a reduced set of nonlinear and linear statistical models (Chapter 5 provides more details in terms of model specifications). Besides, we are also interested in each model's forecasting performance. Therefore, BMA via the marginal likelihoods is applicable.

In other cases, where the number of models under consideration is enormous, implementing BMA via the marginal likelihood would be difficult. One solution to a large K is the Markov Chain Monte Carlo Model Composition (MC^3) algorithm, which has been commonly applied with normal linear regression models to serve the purpose of BMA (see Madigan and York (1995) for MC^3 algorithm, Fernandez, C. et al. (2001) in relation to an investigation of 'benchmark priors' and an application with cross-country growth). Some algorithms, such as reversible jump Markov Chain Monte Carlo method (Green, 1995), the Carlin and Chib method (Carlin and Chib, 1995) and importance sampling method (Clyde, DeSimone and Parmigiani, 1996) are also very popular and can be applied for BMA under different circumstances. For detailed review of the above algorithms, please refer to Carlin and Louis (2000, pp.211-225), and Elliott, Granger and Timmermann (2006, Chapter 10). Clearly, investigating efficient algorithms and developing computational methods of BMA over different kinds of models has been a field with substantial challenges.

2.6 Concluding Remarks

In this chapter, we have described the perspective of Bayesian decision theory along with some computational techniques required for Bayesian inference at a high level of abstraction.

In the Bayesian context, Bayesian inference involve prior elicitations (2. 2), estimations of specified models (2.3), model selections (2.4) and forecasting (2.5). Following this logical modelling approach, we reviewed various methods and computation algorithms that are intensively applied in Chapter 3, 4 and 5.

The logic of the Bayesian modelling approach provides a routine for the rest of the thesis. Future chapters will go through particular models, and show how the reviewed

methods/algorithms are implemented. Thus, the organizations of Chapter 3 and 4 are as the follows: We start by introducing the specified model. Then the likelihood function and elicited priors are presented. We then introduce a posterior simulator for model estimation, followed by a model marginal likelihood estimation method for model comparison. Having developed the analyzing tools, they are evaluated with artificial data and applied with real life data in practical contexts.

Utilizing the tools developed from Chapter 3 and Chapter 4, Chapter 5 focuses on a BMA. More details, such as forecasting evaluations and how the candidate models perform in forecast setting, are revealed in Chapter 5.

Appendix 2.A Calculations of NSE, RNE and CD

All this section is cited from Koop (2003). Before we introduce the calculations for NSE, RNE and CD values, we firstly introduce two theorems in Monte Carlo integrations, which are implied by the central limit theorem,

Theorem 14 *Let $\theta^{(s)}$ for $s = 1, \cdots, S$ be a random sample from $p(\theta \mid y)$, and define*

$$\widehat{g}_S = \frac{1}{S}\sum_{s=1}^{S} g\left(\theta^{(s)}\right)$$

then $\widehat{g}_S$ converges to $E[g(\theta) \mid y]$ as S goes to infinity.

Theorem 15 *Using the step-up and definitions in first Theorem,*

$$\sqrt{S}\left\{\widehat{g}_S - E[g(\theta) \mid y]\right\} \to N\left(0, \sigma_g^2\right)$$

as S goes to infinity, where $\sigma_g^2 = var[g(\theta) \mid y]$.

1. Calculation of NSE

Based on the above two Theorems, we can obtain an estimate of the approximation error in a Monte Carlo integration by using the properties of the Normal distribution. For instance, using the fact that the standard Normal has 95% of its probability located within 1.96 standard deviations from its mean yields the approximate result that

$$\Pr\left[-1.96\frac{\sigma_g}{\sqrt{S}} \leq \widehat{g}_S - E[g(\theta) \mid y] \leq 1.96\frac{\sigma_g}{\sqrt{S}}\right] \approx 0.95$$

and

$$\Pr\left[E[g(\theta) \mid y] - 1.96\frac{\sigma_g}{\sqrt{S}} \leq \widehat{g}_S \leq E[g(\theta) \mid y] + 1.96\frac{\sigma_g}{\sqrt{S}}\right] \approx 0.95$$

Then, the confidence interval for $E\left[g\left(\theta\right) \mid y\right]$ can be approximated by

$$\left[\widehat{g}_S - 1.96\frac{\sigma_g}{\sqrt{S}}, \widehat{g}_S + 1.96\frac{\sigma_g}{\sqrt{S}}\right].$$

By controlling S, we can ensure that $\widehat{g}_S - E\left[g\left(\theta\right) \mid y\right]$ is sufficiently small with a high degree of probability. If enough itertations are taken, NSE should be approximate to 0. In practice, σ_g is unknown, but the Monte Carlo integration procedure allows us to approximate it. The term $\frac{\sigma_g}{\sqrt{S}}$ is known as the Numerical Standard Error (NSE). If $S = 10000$, then the NSE is 1% as big as the posterior standard deviation.

2. Calculation of CD

Geweke (1992) suggests that, based on the intuition that, if a sufficiently large number of draws have been taken, the estimate of $g\left(\theta\right)$ base on the first half of the draws should be essentially the same as the estimate based on the last half. If the two estimates are very different, this indicates either that too few draws have been taken (and estimates are simply inaccurate), or that the effect of initial draw, $\theta^{(0)}$, has not worn off and is contaminating the estimate which uses the first draws.

In the manual book provided by James LeSage, "Applied Econometric using Matlab" (pp.164), the Geweke's convergence diagnostic "represents a test of whether the sample of draws has attained an equilibrium state based on the means of the first 20% of the sample of draws versus the last 50% of the sample. If the Markov chain of draws from the Gibbs sampler has reached an equilibrium state, we would expect the means from these two splits of the sample to be roughly equal. A Z-test of the hypothesis of equality of these two means is carried out. "

Therefore, we could divide the whole Monte Carlo draws into three sets, S_A, S_B and S_c. $\widehat{g}_{S_A}$ and $\widehat{g}_{S_C}$are estimates of $E\left[g\left(\theta\right) \mid y\right]$ using the first S_A replications after discarding the first $S^{(0)}$draws and last S_c replications, respectively. If we define $\frac{\widehat{\sigma_A}}{\sqrt{S_A}}$and $\frac{\widehat{\sigma_C}}{\sqrt{S_C}}$be the NSE of $\widehat{g}_{S_A}$ and $\widehat{g}_{S_C}$. Then

$$CD = \frac{\widehat{g}_{S_A} - \widehat{g}_{S_C}}{\sqrt{\frac{\widehat{\sigma_A}^2}{S_A} + \frac{\widehat{\sigma_C}^2}{S_C}}}$$

According to a central limit theorem,

$$CD \rightarrow N\left(0, 1\right)$$

In Koop (2003, pp.75, line 10), "noting that CD is asymptotically standard normal, a common rule is to conclude that convergence of the MCMC algorithm has occurred if CD is less than 1.96 in absolute value for all parameters"

3. Calculation of RNE

The Relative Numerical Efficiency (RNE) is calculated as follows,

$$RNE = \frac{\sigma_\theta^2}{\sigma_{NW,q}^2}$$

where σ_θ^2 is a direct estimator of the variance, and $\sigma_{NW,q}^2$ is the Newey-West (Newey and West 1987) variance estimator taking the correlation up to lags of $q\%$ of the size of the sample into account. Practical values for q can be 4,8 or even 15%.

Based on the NSE, RNE and CD values, we can diagnose the MCMC convergence sufficiently.

Chapter 3

Evidences of Changing Persistence from A Generalized Stochastic Unit Root Model

3.1 Introduction and Motivations

Application of conventional econometric tests indicates that many macroeconomic time series contain unit roots and are therefore, nonstationary I(1) processes. Some of these results contradict economic theories. Also, there is always an uncertainty about shifts in the persistence of a time series, such as shifts from stationarity $I(0)$ to nonstationarity $I(1)$ or vice versa. Based on these unsolved problems in applications of time series tests, this chapter focuses on two aspects: One is about analyzing the uncertainty of changing persistence in the time series, with the purpose of solving the Purchasing Power Parity puzzle; the other is about analyzing the uncertainty of deterministic trend in the time series. How sure one can be that an economic time series has a deterministic trend, especially when there is a change in the underlying properties? This chapter makes use of Bayesian model estimation and model comparison techniques to investigate the deterministic time trend and changes in persistence[1] in time series with a Generalized Stochastic Unit Root (GSTUR hereafter) model. This coefficient nonlinear GSTUR model is specified to allow for shifts in the values of the roots in a time series. Empirical analysis indicates that the GSTUR model could provide new insights in time series studies.

[1] *A change in persistence* in this Chapter specially refers to the root of the series jump to/above the unity, or vice versa.

To motivate the GSTUR model, first, we briefly review the Purchasing Power Parity (PPP hereafter) puzzle in the literature. Second, we address the mystery of the deterministic trend in time series modelling. Finally, we introduce the properties of GSTUR model and present why the GSTUR model could potentially solve these puzzles.

3.1.1 A Puzzle of Purchasing Power Parity

Understanding the underlying properties of the real exchange rate is considered vital as well as is PPP[2], which is a cornerstone of many open-economy macroeconomic models. Further to this, policy makers have to consult the movements of real exchange rates in decisions of settings of exchange rate parity, evaluating the stability of trading flows and determining the degree of misalignment of the nominal exchange rate. The PPP puzzle is that the logarithm real Exchange Rate (ER hereafter)[3] series exhibits high persistence in empirical findings. In the survey of PPP literature (see Rogoff (1996), Taylor (1995)), problem are raised such that it takes a long time for the real exchange to revert to the stable equilibrium level.

Early investigations of PPP were based on linear models, utilizing the conventional augmented Dickey–Fuller (ADF) tests for the real exchange rate series, and results turned out to be invalidity of the PPP. The reason for a failure to reject the unit-root could be due to our lack of competency in the developed statistics knowledge that the unit root tests and cointegration tests have low power. For details of low power of the conventional ADF test in distinguishing the unit root and near unit root behaviour, refer to Lothian and Taylor (1997). Then, researchers have to resort to long-spans of data (e.g. Kim (1990); Lothian and Taylor (1996)), or panel unit root studies (e.g. Frankel and Rose (1996); Papell (1997), O'Connell (1998); Taylor and Sarno (1998)), or time-series models that account for the possibility of nonlinear mean reversion towards PPP (e.g. Obstfeld and Taylor (1997), Baum et al. (2001), Taylor

[2]PPP is a propostion that the national price levels should be equal when expressed in a common currency.

[3]Real Exchange Rate, which can be interpreted as a measuremnet of the deviation from the absolute PPP, is calculated as the norminal exchange rates adjusted by the price levels

$$RER = NER + p^* - p$$

where RER denotes the real exchange rate between the foreign currency and domestic currency. NER denotes the nominal dollar-sterling exchange rate, which is the amount of US dollar to pay for 1 UK Pound Sterling. p^* denotes the UK price level and p denotes the US price level. The data used in this thesis is the same data used in Engel and Kim (1999), monthly data from 1885:1 to 1995:2, which was originally provided by Graciela Kaminsky.

et al. (2001)). These researches have the purpose of overcoming the conventional low testing power problem and trying to find stationarity in the real exchange rates and the validity of PPP (see Sarno 2005 for a recent survey). In a very recent strand, nonlinear models such as the Threshold Autoregressive (TAR model applied by Sarno, Taylor and Chowdhury 2004) and STAR (Smooth Transition Autoregressive applied by Sarantis 1999), Exponential Smooth Autoregressive (ESTAR applied by Michael, Nobay and Peel 1997, Baum, Barkoulas and Caglayan 2001), Self-Exciting Threshold Autoregressive (SETAR applied in a very recent paper by Serletis and Shahmoradi, 2007) for the PPP puzzle, are motivated by the fact that a transaction cost could be the reason for nonlinear adjustment between the price differentiates and nominal exchange rates.

Given the large literature, regarding the PPP puzzle, this Chapter is motived by the following questions: (1) With what kind of models, can we export the real exchange rates' underlying properties, in order to find support for the PPP theory? (2) Further to this, we want to know whether monetary events are in correspondence with the shifts in the structural dynamics of PPP deviations.

3.1.2 An Uncertainty Associated with Deterministic Time Trends

In time series modelling, a common question is asked: Why should we worry about integration degrees of a time series, and why should we care about the characteristics (stochastic and/or deterministic) of the trends in the first place[4]? One answer to the above questions is: analyzing those properties, such as integration degree and the existence of a deterministic time trend, are crucial because those properties will determine the ultimate forecasts.

However, it is difficult to identify a deterministic time trend in a sense that a Trend Stationary (TS) series and a Difference Stationary (DS) series can hardly be distinguished from each other. In empirical studies, the results of testing whether the macroeconomic time series are TS or DS are full of ambiguity (see Nelson and Plosser (1982), DeJong and Whiteman (1991a, b), Phillips (1991), Kwiatkowski, Phillips, Schmidt, and Shin (1992)).

Furthermore, the integration degree of a time series might change over time. With applications of the U.S. macroeconomic data sets, a lot of evidence was found concerning the changes in persistence (Kim 2000, Kim and Amador 2002, Busetti and

[4]A stochastically trending variable refers to a variable that is integrated with order 1

Taylor 2004, Harvey et al. 2006, Cogley, Primiceri and Sargent 2008). Therefore, it is sensible to argue that a deterministic time trend hypothesis is often rejected because of the changes in the persistence. Some crucial questions appear immediately, such as how sure are we that economic time series have a deterministic trend if the underlying properties changed and whether the variations of persistence correspond with historical events?

More explicitly, in regards to the uncertainties of deterministic trends and changing persistence, we also would like to ask the following question: can we find a flexible model to accommodate a possible deterministic linear time trend when the underlying property changes?

3.1.3 Solutions to the Puzzles: GSTUR Model with Bayesian inference

GSTUR models provide a potential solution to the questions raised in sections 3.1.1 and 3.1.2. In a recent strand of time series modelling, nonlinear models, such as TAR (Caner and Hansen 2001), STAR (van Dijk et al. 2002), ESTAR (Kapetanios et al. 2003), and alternative forms of stationarity, such as ARFIMA (Koop et al. 1997), have been proposed for reconciling the nonstationarity in macroeconomic time series to economic theory. The GSTUR model is a class of nonlinear models, yet it has not been fully investigated.

Granger and Swanson (1997) proposed a highly parameterized model, the Stochastic Unit Root (STUR) model

$$y_t = \exp(\alpha_t) y_{t-1} + \varepsilon_t \tag{3.1}$$

$$\alpha_t = \phi_0 + \phi_1 \alpha_{t-1} + \eta_t \tag{3.2}$$

$$E(\alpha_t) = 0$$

where ε_t is $i.i.d.N(0, \sigma_\varepsilon^2)$ and η_t is $i.i.d.N\left(0, \sigma_\eta^2\right)$. One main distinctive feature of the STUR model is that it allows for the persistence of macroeconomic series to vary with time. This changing persistence property could be a characteristic of a series that cannot be differenced to stationarity. The GSTUR model is a flexible model in capturing the statistical properties of the data which are attractive. It does not have economic implications as does STAR. However, the STUR process may shed light on the data properties on the levels as pointed out in Granger and Swanson (1997).

Although the STUR model has many appealing properties, it has not yet been fully studied and widely applied in the literature.

The literature with regards to the STUR model consists of the following: Granger and Swanson (1997) estimate the STUR model and evaluated its forecasting performance. Testing a STUR process is developed and evaluated in Taylor and van Dijk (2002). Some empirical applications involved the STUR model can be found in Bleaney, Leybourne and Mizen (1999) for real ER, Sollis, Leybourne and Newbold (2000) for worldwide stock market indices, Yoon (2005) with the risk-free interest rate. Regarding the properties of a STUR process, Yoon (2006) provides intensive reviews. To fully understand the STUR process in (3.1 and 3.2), and its potential with empirical applications, we summarize them as follows based on Yoon (2006):

1. y_t has a unit root only on average and can be explosive for some periods. However, y_t is not a non-stationary process.

2. The STUR processes have very heavy tails and extreme values are very likely to occur.

3. A STUR process cannot be transformed to stationarity by taking differences.

Because of property 1, STUR processes are hard to distinguish from a Random Walk (RW) process[5], especially when $var\left(\alpha_t\right)$ is small. Therefore, "taking differences cannot make the STUR processes weakly stationary" (Yoon, 2006).

Within the understanding of the STUR process, we generalized the STUR model to a Generalized STUR (GSTUR) model. The generalization is carried out by assuming the existence of a deterministic time trend and a shift in the underlying process of y_t:

$$\nu_t = y_t - \gamma - \delta t \tag{3.3}$$

The motivation for the generalization is based on the fact that there is always an uncertainty as to whether a macroeconomic series is trend stationary (TS) or difference stationary (DS), or neither, using the conventional econometric tests. In the GSTUR model, ν_t is a STUR process.

$$\nu_t = \exp(\alpha_t)\nu_{t-1} + \sum_{i=1}^{l}\lambda_i \triangle \nu_{t-i} + \varepsilon_t \tag{3.4}$$

$$\alpha_t = \phi_0 + \phi_1\alpha_{t-1} + \cdots + \phi_p\alpha_{t-p} + \eta_t \tag{3.5}$$

[5]Random Walk process: $y_t = y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

where ε_t is $i.i.d.N(0, \sigma_\varepsilon^2)$ and η_t is $i.i.d.N\left(0, \sigma_\eta^2\right)$.

Standard algebra shows that this GSTUR model has the following desirable features: (1).The changes in persistence could have taken place at any time point. (2).The values of the parameters characterize a new degree of persistence, which is potentially dependent on its own lagged values. (3).Previous information and the previous degree of persistence provide information of newly commencing information. (4).A deterministic trend might exist regardless the variations of persistence degrees.

While modelling the changes in persistence as a stochastic process seems attractive in both the STUR and GSTUR models, estimation was problematic. Granger and Swanson (1997) used the approximate maximum likelihood (approximate ML hereafter) method (Guyton et al.1986) and the approximate minimum sum squares method to estimate the parameters in a STUR model. These estimation methods produced 'fairly imprecise' and 'wild estimates' (see Granger and Swanson, 1997). The inefficient estimates using approximate ML estimation may be due to the idiosyncratic properties of the STUR process. Yoon (2006) states, since "the STUR process do not have any finite moments as $t \to \infty$ ", the approximate ML is not applicable because " the approximate ML method requires the existence of moments of high enough order for the estimates to be consistent and asymptotically normal". To solve the problem of estimating the highly parameterized STUR model, Jones and Marriott (1999), JM hereafter, provided a Bayesian method for parameter estimations in (3.1, and 3.2). For details of estimations in Bayesian framework, please refer to Chapter 1.3.2 (Bayesian estimations).

In this chapter, Bayesian analysis of a GSTUR model (in 3.4 and 3.5) is presented. The motivation for using the Bayesian method are as follows: (1) From the model estimation aspect, high dimensional model with latent variables can be estimated with Bayesian MCMC methods easily. (2) From the model comparison aspects, marginal likelihoods in Bayesian model selection procedures could not only tell which proposition is the most supported, but also to what extent the proposition is favoured according to the data information. Besides, a comparison between the nonlinear GSTUR model and the linear RW model can be achieved via the Bayes factors, which is not available in the classical framework .

The remainder of this chapter is organized as follows: Section 2 develops a pack of tools for practitioners, who wish to apply the GSTUR model with empirical applications. This completed tool pack includes a sampler for estimation and the GSTUR model's marginal likelihood evaluation in a Bayesian framework. In Section 3, the tool pack is evaluated with a simulated artificial data series. Section 4 presents the

findings using the GSTUR model with applications of one of the Nelson and Plosser's data series, the S&P 500 series. The U.K. /U.S. long run exchange rates data is also applied for empirical illustrations. Section 5 contains brief concluding remarks.

3.2 Bayesian inference

According to the GSTUR model (3.3, 3.4 and 3.5), α_t is literarily an AR(P) process. The roots of the polynomials α_t are restricted to a region within the unit circle. Then an unconditional mean μ_α of the stationary α_t process has the following expression:

$$\mu_\alpha = \frac{\phi_0}{1 - \sum_{j=1}^{p} \phi_j}$$

also,

$$\rho_t = \exp(\alpha_t) \tag{3.6}$$

We begin by introducing some notations:

In this GSTUR model, $F_t = (y_{1-l}, \cdots, y_1, \cdots, y_t)'$. The whole sample of observations is denoted as $y = (y_{1-l}, \cdots, y_1, \cdots, y_n)'$ with a sample size of $T = n + l$, where the first $l + 1$ values $y_{init} = (y_{1-l}, \cdots, y_1)'$ are treated as the starting values. Latent variable α, where $\alpha = (\alpha_2, \cdots, \alpha_n)'$, denotes the series of stochastic roots over the whole time T period, which are structurally unobservable and indicate changes in the persistence of y. The initial state of α is denoted as $\alpha_{init} = (\alpha_{2-p}, \cdots, \alpha_1)'$, in which all elements are equal to 0. Also, we define α_{-t} as the lagged data sequence of α_t, where $\alpha_{-t} = (\alpha_{2-p}, ..., \alpha_{t-1})'$

Vector ϕ and λ are defined as $\phi = \left(\phi_1, \cdots, \phi_p\right)'$, $\lambda = (\lambda_1, \cdots, \lambda_l)'$. According to equation (3.6), vector $\rho = (\rho_2, \cdots, \rho_n)'$ associate with α. The error precisions are denoted as $h_\varepsilon = \sigma_\varepsilon^{-2}$ and $h_\eta = \sigma_\eta^{-2}$. θ stands for a vector of all the parameters of interest, where $\theta = (\Omega, \lambda, \phi, \mu_\alpha, h_\varepsilon, h_\eta)$, where $\Omega := (\gamma, \delta)$.

As motivation, we should mention that ρ_t, where $t = 2, \cdots, n$, varies stochastically in the GSTUR process. To investigate if a process y undergoes shifts in persistence, or being parameter nonlinear, we can make inference from the estimates of ρ_t. According to equation (3.5), a small σ_η^2, which is the variance of η_t in the transition equation, enables α_t to vary around the conditional mean. Consequently, process y may maintain stationarity ($I(0)$) if $\rho_t < 1$, but becomes a process with higher persistence ($I(1)$) if the averaged values of $\rho_t > 1$. Therefore, the STUR process is not a pure $I(1)$ or a pure $I(0)$ process, but follows a near integrated process. These idiosyncratic properties

of a STUR process make it very difficult to distinguish from a RW process .

How frequently the process becomes temporarily nonstationary depends upon how negative μ_α is and the size of σ_η^2. If $\mu_\alpha = 0$, the stochastic root ρ_t will vary around the unity such that the process is as often stationary as it is nonstationary. The difference stationary RW process is nested within the GSTUR process at the point where $\mu_\alpha = 0$ and $\sigma_\eta^2 = 0$ such that ρ_t will be a constant and equal to 1 for all the t. Therefore, the behaviour of a process could be learnt about by observing the distribution of μ_α and the evolution of ρ_t over time. Hence, our estimations focus on μ_α, ρ_t, and δ, where δ represent a deterministic time trend behavior in y.

3.2.1 Bayesian Model Estimation

The parameters that have to be estimated in the GSTUR model are θ and α. We first develop the forms of prior, likelihood and the full posterior conditionals. Then, the Gibbs sampling algorithm and Metropolis-Hastings algorithm are implemented for the parameter estimations (for reviews of these algorithms, please see Chapter 2.3.2).

As θ denotes a vector, which includes all the parameters of interest in the GSTUR model, the properties of vector θ will determine the properties of the latent data α. As a standard practice with Bayesian methods, the latent data in α are treated as 'pseudo' parameters whose behaviour is governed by values of the elements of θ, in particular the values of μ_α, σ_η^2 , ϕ and the initial values α_{init}. Treating α as data, the vector α enters the posterior via the likelihood. As the α are both stochastic and unobserved then they must be estimated as must the parameters in θ. Thus, Bayesian analysis uses the joint density of θ, and α conditional on y will be

$$\begin{aligned} p(\theta, \alpha|y) &\propto p(y|\alpha, \theta)\, p(\alpha, \theta) \\ &= p(y \mid \theta, \alpha)\, p(\alpha \mid \theta)\, p(\theta) \end{aligned}$$

According to Equation (3.3), (3.5) and (3.4), the likelihood function of $p(y \mid \alpha, \theta)$

is

$$
\begin{aligned}
p\left(y \mid \alpha, \theta\right) &= \prod_{t=2}^{n} p\left(y_t \mid \theta, \alpha_t, F_{t-1}\right)^{6} \\
&= \frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{\frac{n-1}{2}}} \exp\left\{ \begin{array}{c} -\frac{1}{2\sigma_{\varepsilon}^{2}} \sum_{t=2}^{n} \\ \left(\begin{array}{c} y_t - \exp(\alpha_t) y_{t-1} - \sum_{i=1}^{l} \lambda_i \triangle y_{t-i} - \\ \gamma\left[1 - \exp(\alpha_t)\right] - \\ \delta\left[t - \exp(\alpha_t) t + \exp(\alpha_t) - \sum_{i=1}^{l} \lambda_i\right] \end{array} \right)^{2} \end{array} \right\}
\end{aligned}
$$

The joint density of the α is given by the sequence of conditionals as

$$
\begin{aligned}
p\left(\alpha \mid \theta\right) &= p\left(\alpha_n, \alpha_{n-1}, ..., \alpha_{2-p} \mid \theta\right) \\
&= p\left(\alpha_n \mid \alpha_{n-1}, ..., \alpha_{2-p}, \theta\right) p\left(\alpha_{n-1} \mid \alpha_{n-2}, ..., \alpha_{2-p}, \theta\right) ... \\
&\quad p\left(\alpha_2 \mid \alpha_1, ..., \alpha_{2-p}, \theta\right) p\left(\alpha_1, ..., \alpha_{2-p} \mid \theta\right) \\
&= \prod_{t=2}^{n} p\left(\alpha_t \mid \alpha_{t-1}, ..., \alpha_{2-p}, \theta\right)
\end{aligned}
$$

where the likelihood $p\left(y, \alpha|\theta\right)$ equals $p\left(y|\alpha, \theta\right) p\left(\alpha|\theta\right)$.

An important issue in calculating Bayes factors is Bartlett's paradox (Bartlett, 1957). This paradox implies that improper priors cannot be used if one intends to compute informative Bayes factors. The issue is that in this case the model with the smallest dimension will be selected with probability one regardless of the information in the data. Thus we confine ourselves to using proper priors and generally follow the recommendations of JM with some important exceptions. We assume that each parameter is a "priori" independent and we formulate prior distributions on the parameters of interest. The prior of θ, which include the variables of interest takes the following form:

$$
\begin{aligned}
p\left(\theta\right) &= p\left(\Omega, \lambda, \phi, \mu_{\alpha}, h_{\varepsilon}, h_{\eta}\right) \\
&= p\left(\mu_{\alpha}\right) p\left(\gamma\right) \cdots p\left(\phi_p\right) p\left(h_{\varepsilon}\right) p\left(h_{\eta}\right)
\end{aligned} \tag{3.7}
$$

The notations of relevant distributions are given as follows: The prior distributions are chosen as Normal for μ_{α}, denoted as $f_N\left(\underline{\mu}_{\alpha}, \underline{V}_{\alpha}\right)$ and Multivariate Normal for $\{\Omega\}$ and $\{\lambda\}$, which are denoted as $f_{MN}\left(\underline{\mu}_{\Omega}, \underline{V}_{\Omega}\right)$ and $f_{MN}\left(\underline{\mu}_{\lambda}, \underline{V}_{\lambda}\right)$. The prior of ϕ_i is $\phi_i \sim f_N\left(\underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}\right) 1\left(\|z_j\| > 1\right)$. Since α_t is restricted to be a stationary $AR(p)$

process, the roots of the estimated polynomial (also in Enders 1995, pp.96):

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

should all lie outside of the unit circle. In the prior, $1(\|z_j\| > 1)$ is an indicator function for the event that $\phi_i : i = 1, \cdots, p$ jointly satisfy the stationary condition. The error precisions' priors are chosen as $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ and $h_\eta \sim f_{\Gamma\eta}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$ [7].

Since the priors represent our "priori" belief over the parameters, selecting the values in the prior distributions are important. The selected values in Table (3.1) represents our belief that μ_α varies around $\ln 0.9$ with a small variance $\underline{V}_\alpha$. In JM, the prior of μ_α is chosen as normal distribution conditional on σ_η^2 such that $\mu_\alpha \sim f_N(0, \sigma_\eta^2)$. Besides, in JM, the selected priors for the σ_ε^2 and σ_η^2 indicate $0 < \sigma_\varepsilon^2 < \infty$ and $0 < \sigma_\eta^2 < \infty$ in corresponding to "little prior information", denoted as "diffuse" priors (see JM for prior elicitations and Zellner, 1971 for details regarding diffuse priors). In this chapter, we assume μ_α and σ_η^2 are independent. More over, rather than choosing diffused priors for σ_ε^2 and σ_η^2, we elicit a prior in a belief that σ_η^2 is a small value, which corresponds to the properties of a STUR process such that the values of σ_η^2 should be small, which may allow the root to vary above 1 after certain periods of time, but not diverge far from the previous period's root. For the purpose of empirical applications with a GSTUR model, the probability of σ_ε^2 being extremely large should be small.

Table 3.1: Summarized Prior Properties in GSTUR

parameters	Selected Values in the Prior			
μ_α	$\underline{\mu}_\alpha$	$\ln 0.9$	$\underline{V}_\alpha$	0.1^2
ϕ_i	$\underline{\mu}_{\phi_i}$	1	$\underline{V}_{\phi_i}$	0.1
h_ε	$\underline{\alpha}_\varepsilon$	1.1	$\underline{\beta}_\varepsilon$	0.2
h_η	$\underline{\alpha}_\eta$	1.5	$\underline{\beta}_\eta$	2.5
Ω	$\underline{\mu}_\Omega$	$\binom{0}{0}$	$\underline{V}_\Omega$	$10^4 eye(2)$
λ	$\underline{\mu}_\lambda$	$(0, \cdots, 0)'$	$\underline{V}_\lambda$	$10^4 eye(i)$

Figure (3-2) plots the prior of ϕ_1 when $p = 1$. This prior indicates that the α_t and α_{t-1} has positive correlation and it tends to be high.

[7] Refer to Appendix 3.A for the prior densities.

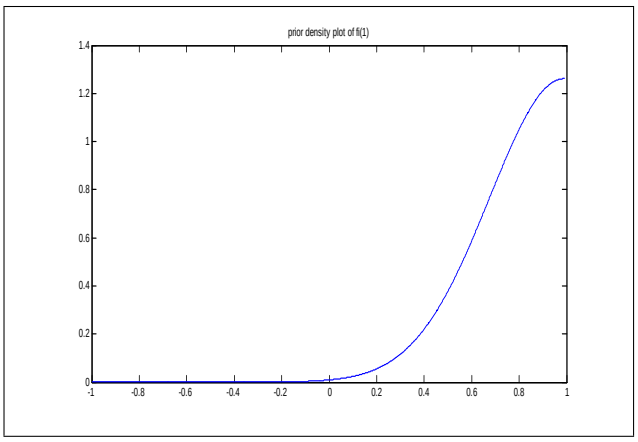


Figure 3-1: Plot of fi prior p=1

In Table (3.1), the selected prior for h_ε indicates the error precision σ_ε^{-2} has high densities around small values (see Fig 3-2 for the density plot of h_ε). Consequently, the σ_ε^2 distribution, in which σ_ε^2 is the inverse of h_ε, has a wide scale (see Fig 3-3 for the density plot of σ_ε^2). Similarly, the selected values of $\underline{\alpha}_\eta$ and $\underline{\beta}_\eta$ in the prior for h_η indicates a small value of σ_η^2 in a narrow scale (see Fig 3-4 for the density plot of error precision h_η and the corresponding density plot of σ_η^2 in Fig 3-5)[8].

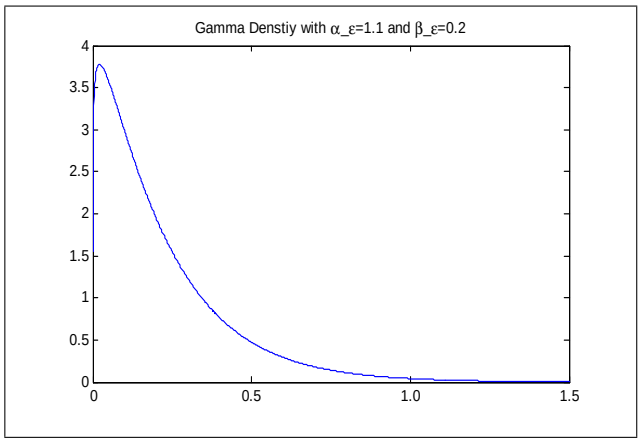


Figure 3-2: Prior *pdf* for Error Precision in GSTUR $\sigma_\varepsilon^{-2} \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$

Given the above prior information, we are able to derive the conditional posterior distributions that can be used in the Gibbs sampling scheme. The full conditional density of most of the parameters have standard forms. Via the Bayes Theorem, the joint posterior density for (θ, α) is then

[8]For the relations between Gamma and inverse Gamma distributions, refer to Appendix B.

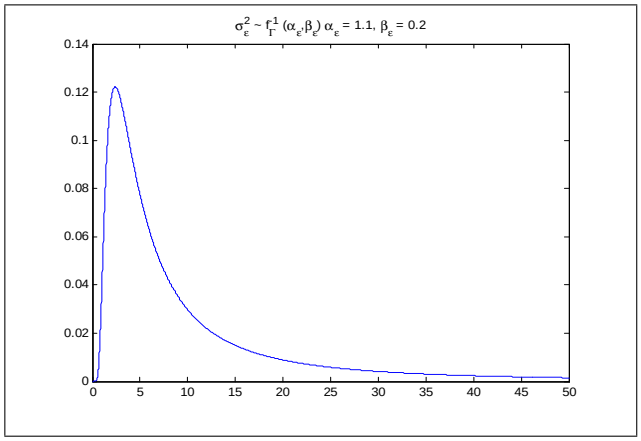


Figure 3-3: Prior *pdf* for Error in GSTUR $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$

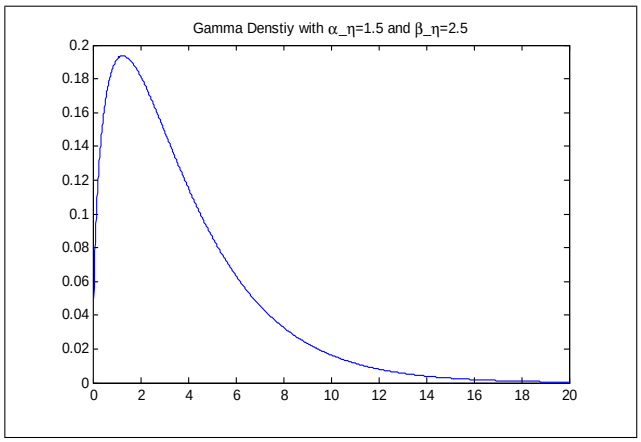


Figure 3-4: Prior *pdf* for Error Precision in GSTUR $\sigma_\eta^{-2} \sim f_{\Gamma_\eta}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$

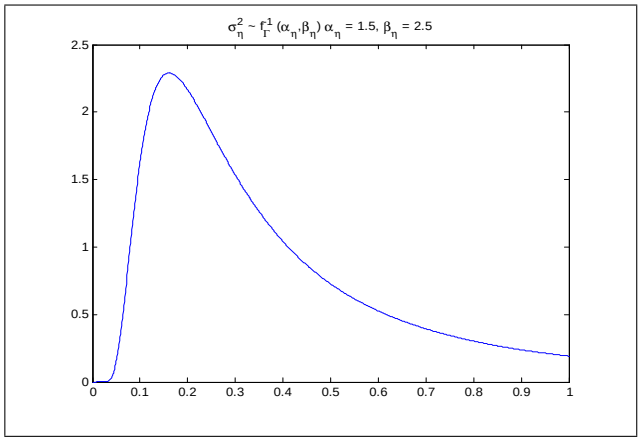


Figure 3-5: Prior *pdf* for Error in GSTUR $\sigma_\eta^2 \sim f_{\Gamma_\eta}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$

$$\begin{aligned} p(\theta, \alpha \mid y) &\propto \prod_{t=2}^{n} p(y_t \mid \alpha, \theta, F_{t-1})\, p(\alpha, \theta) \\ &= \prod_{t=2}^{n} p(y_t \mid \alpha, \theta, F_{t-1})\, p(\alpha \mid \theta)\, p(\theta) \end{aligned} \tag{3.8}$$

Using many of the results from JM, we can present the following conditional distributions: the posterior conditionals of $\{\Omega\} \sim f_{MN}\left(\overline{\mu}_\Omega, \overline{V}_\Omega\right)$, $\{\lambda\} \sim f_{MN}\left(\overline{\mu}_\lambda, \overline{V}_\lambda\right)$, $\mu_\alpha \sim f_N\left(\overline{\mu}_\alpha, \overline{V}_{\mu_\alpha}\right)$, $\phi_i \sim f_N\left(\overline{\mu}_{\phi_i}, \overline{V}_{\phi_i}\right) 1\left(\|z_j\| > 1\right)$, where $\phi_i : i = 1, \cdots, p$ jointly satisfy the stationary condition. Finally, the posterior conditionals of error precisions are $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right)$ and $h_\eta \sim f_\Gamma\left(\overline{\alpha}_\eta, \overline{\beta}_\eta\right)$ [9] and the posterior conditional for α_t is described as follows:

$$\begin{aligned} &p(\alpha_t \mid \Omega, h_\varepsilon, h_\eta, \lambda, \mu_\alpha, \phi, \alpha_{-t}, y) \\ \propto\ &\exp\left[-\frac{\nu_{t-1}^2}{2\sigma_\varepsilon^2}\left(e^{\alpha_t} - \frac{\nu_t - \sum_{i=1}^{l}\lambda_i \triangle \nu_{t-i}}{\nu_{t-1}}\right)^2 - \frac{\vartheta(\alpha_t)}{2\sigma_\eta^2}\left(\alpha_t - \frac{\tau(\alpha_t)}{\vartheta(\alpha_t)}\right)^2\right] \end{aligned}$$

where $\nu_t = y_t - \gamma - \delta t$, $\vartheta(\phi)$ is a function of ϕ, and $\tau(\alpha_t)$is a function of $\mu_\alpha, \phi, \alpha_{-t}, p$ and t.

The latent variable α_t follows a non-standard conditional distribution. JM applied

[9] Refer to Appendix 3.B for the derivations of the posterior conditionals.

a ratio of uniforms method(see Devroye, 1986) to sample α_t. As we can see that the posterior conditionals of α_t is similar to a normal density, the Metropolis-Hastings (M-H) algorithm (for details of M-H algorithm, please refer to Chapter 2.3.3) can be implemented to simulate α_t. The proposed M-H algorithm exhibits speedy convergence in later experiments, which is an efficient sampling scheme for α. The full expression of $\vartheta(\phi)$ and $\tau(\alpha_t)$ are provided in Appendix 3.B. As all the posterior conditionals are attainable, we apply this following MCMC procedure to simulate the posterior distributions for the parameters of interest in the GSTUR model.

Algorithm 1: Posterior Simulator of GSTUR-Implementations of Gibbs and MH Sampling Algorithms

1. Give initial values to $\sigma_\varepsilon^2, \sigma_\eta^2, \mu_\alpha, \phi, \boldsymbol{\alpha}, \lambda, \Omega$
2. Sample σ_ε^2 from $\left(\sigma_\varepsilon^2 \mid y, \Omega, \lambda, \alpha, \phi, \mu_\alpha, \sigma_\eta^2\right) \sim f_{\Gamma_\varepsilon}^{-1}\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right)$
3. Sample σ_η^2 from $\left(\sigma_\eta^2 \mid y, \Omega, \lambda, \alpha, \phi, \mu_\alpha, \sigma_\varepsilon^2\right) \sim f_{\Gamma_\eta}^{-1}\left(\overline{\alpha}_\eta, \overline{\beta}_\eta\right)$
4. Sample μ_α from $\left(\mu_\alpha \mid y, \Omega, \lambda, \alpha, \phi, \sigma_\varepsilon^2, \sigma_\eta^2\right) \sim f_N\left(\overline{\mu}_\alpha, \overline{V}_{\mu_\alpha}\right)$
5. Sample ϕ_i from $\left(\phi_i \mid y, \Omega, \lambda, \alpha, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2\right) \sim f_N\left(\overline{\mu}_{\phi_i}, \overline{V}_{\phi_i}\right)$ draw ϕ_i, if all $\phi_i s$ jointly satisfy the stationary condition$\|z_j\| > 1$, accept the vector of ϕ draw, otherwise draw the whole vector ϕ again.
6. Sample $\boldsymbol{\alpha}$ from $p\left(\alpha_t \mid y, \Omega, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2, \alpha_{-t}\right)$ using independent chain M-H algorithm, in which a univariate t-denstiy is chosen as the candidate generating density.
7. Sample λ from $\lambda \mid y, \Omega, \alpha, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2 \sim f_{MN}\left(\overline{\mu}_\lambda, \overline{V}_\lambda\right)$
8. Sample Ω from $\Omega \mid y, \alpha, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2 \sim f_{MN}\left(\overline{\mu}_\Omega, \overline{V}_\Omega\right)$
9. Goto 2
10. Jump out of the loop until S iterations are accomplished.

3.2.2 Bayesian Model Comparison

In this chapter, one of the main concerns is the existence of a deterministic time trend such that δ equals 0. Another concern is the parameter nonlinearity in the process such that ρ is time-variant. Thus, a group of models holding different hypotheses will be compared. According to each model's probability based on the available data

information, the optimal model to represent the data will be selected. If we weight average the models according to the model probabilities, uncertainties could be reduced. This model selection procedure is easily achievable using the Bayes Factors even if entertained models are not nested. Please refer to Table (2.1) in chapter 2.4 for the criterion of model selection using Bayes Factors.

As the GSTUR model is highly dimensional, the non-linearity in parameters and/or the choice of nonconjugate priors that makes the marginal likelihood unavailable in closed form. The model is highly dimensional. Therefore, the marginal likelihood cannot be obtained using numerical integration method in a straight forward manner, because "the brute-force integration method suffers from the curse of dimensionality" (Bos, 2002). In the case where a Gibbs Sampler is already implemented in section 3.2.1 and all the full posterior conditionals are known, it is possible to approximate the marginal likelihood of each competing model from the posterior samples using the approach introduced by Chib (1995). In chapter 2.4, we provided a brief review of using the Chib method to compute Bayes factors. In this chapter, Chib method is applied to compute the marginal likelihoods.

The idea of the Chib method starts from the basic 'marginal likelihood identity'. If we denote $\underline{C}$ and $\overline{C}$ as the integrating constants in the truncated normal priors and posteriors, in Chib (1995), the marginal density of $y = (y_2, ..., y_n)'$, can be written as:

$$p(y) = \frac{\prod_{t=2}^{n} p(y_t \mid \theta) p(\theta) / \underline{C}}{p(\theta \mid y) / \overline{C}}$$

where the numerator is the product of the sampling density and the prior with all integrating constants included, and the denominator is the posterior density of θ including the integrating constant. In the transition equation, as ϕ is truncated to satisfy the stationary restriction, the integrating constant for the ϕ prior and the ϕ posterior, denoted as $\underline{C}_\phi$ and $\overline{C}_\phi$ [10]respectively, can be evaluated in a simulation manner (Refer to Judge et al. (1985, pp.128) for details). As the above 'marginal likelihood identity' holds for any θ, say θ^*, the log marginal likelihood can be approximated as

$$\ln \widehat{p}(y) \propto \sum_{t=2}^{n} \ln \widehat{p}(y_t \mid \theta^*, F_{t-1}) + \ln \left[\widehat{p}(\theta^*) / \underline{C}_\phi\right] - \ln \left[\widehat{p}(\theta^* \mid y) / \overline{C}_\phi\right] \tag{3.9}$$

where $\widehat{p}(\theta^*)$ is available straight forward by evaluating the prior densities at θ^*[11]. The

[10]The estimation of $\overline{C}_\phi$ used the method explained in Koop (2003, pp.134).

[11]For the approximation of prior ordinates evaluated as $p(\theta^*) = p\left(\phi^*, \mu_\alpha^*, h_\eta^*, \gamma^* \delta^*, \lambda^*, h_\varepsilon^*\right)$ please

posterior ordinates $\ln \widehat{p}(\theta^* \mid y)$ and the likelihood ordinates $\ln \widehat{p}(y_t \mid \theta^*, F_{t-1})$ at θ^* can be evaluated using the Monte Carlo samples. θ^* is normally chosen as the posterior means or modes.

To estimate the posterior ordinate $\widehat{p}(\theta^* \mid y) = \widehat{p}(\phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, \lambda^*, h_\varepsilon^* \mid y)$, we may use a marginal/conditional decomposition together with the outputs from the original and subsequent 'reduced MCMC runs', or so called 'replicated Gibbs sampler procedures' (see chapter 2.4 for details).

Using a multiplication rule

$$\begin{aligned}\widehat{p}(\phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, \lambda^*, h_\varepsilon^* \mid y) &\propto \widehat{p}(\phi^* \mid y) \cdot \widehat{p}\,(\mu_\alpha^* \mid y, \phi^*) \\ &\quad \cdot \widehat{p}\left(h_\eta^* \mid y, \phi^*, \mu_\alpha^*\right) \cdot \widehat{p}\left(\gamma^*, \delta^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*\right) \\ \cdot \widehat{p}(\lambda^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*) &\cdot \widehat{p}(h_\varepsilon^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, \lambda^*) \quad (3.10)\end{aligned}$$

The first term $\widehat{p}\,(\phi^* \mid y)$ in Equation (3.10) can be estimated from the Gibbs sampler outputs as

$$\widehat{p}\,(\phi^* \mid y) \simeq \frac{1}{S}\sum_{s=1}^{S} p(\phi^* \mid y, \mu_\alpha^{(s)}, h_\eta^{(s)}, \Omega^{(s)}, \lambda^{(s)}, h_\varepsilon^{(s)})/\overline{C}_\phi$$

where $\overline{C}_\phi$ is the proportion of accepted draws to the total number of draws. Evaluations for $\widehat{p}\,(\mu_\alpha^* \mid y, \phi^*)$ is taken using a fixed ϕ, where $\phi = \phi^*$ and ϕ^* is chosen as the posterior mean of ϕ.

$$\widehat{p}\,(\mu_\alpha^* \mid y, \phi^*) \simeq \frac{1}{S}\sum_{s=1}^{S} p(\mu_\alpha^* \mid y, \phi^*, h_\eta^{(s)}, \Omega^{(s)}, \lambda^{(s)}, h_\varepsilon^{(s)})$$

and

$$\widehat{p}\left(h_\eta^* \mid y, \phi^*, \mu_\alpha^*\right) \simeq \frac{1}{S}\sum_{s=1}^{S} p(h_\eta^* \mid y, \phi^*, \mu_\alpha^*, \Omega^{(s)}, \lambda^{(s)}, h_\varepsilon^{(s)})$$

$$\widehat{p}\left(\gamma^*\delta^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*\right) \simeq \frac{1}{S}\sum_{s=1}^{S} p(\gamma^*\delta^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \lambda^{(s)}, h_\varepsilon^{(s)})$$

$$\widehat{p}(\lambda^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*) \simeq \frac{1}{S}\sum_{s=1}^{S} p(\lambda^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, h_\varepsilon^{(s)})$$

$$\widehat{p}(h_\varepsilon^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, \lambda^*) = p\left(h_\varepsilon^* \mid y, \phi^*, \mu_\alpha^*, h_\eta^*, \gamma^*\delta^*, \lambda^*\right)$$

With a total of 4 iterated Gibbs runs, we are able to estimate the posterior ordinates.

With the prior ordinates and the posterior ordinates available in a straight forward manner, the likelihood ordinates $p(y \mid \theta^*)$ involves an application of the Auxiliary

refer to Appendix 3.A for the forms of the prior densities.

Particle Filter (APF).

Since y_t is a function of latent variable α_t, conditional on $F_{t-1} = (y_{1-l}, \cdots, y_1, \cdots, y_{t-1})$, the estimation of the likelihood ordinate $p(y_t \mid \theta^*, F_{t-1})$ evolves an integration of the joint likelihood $p(y_t, \alpha_t \mid \theta^*, F_{t-1})$, since

$$\begin{aligned} p\left(y_t \mid \theta^*, F_{t-1}\right) &= \int p(y_t, \alpha_t \mid \theta^*, F_{t-1}) d\alpha_t \\ &= \int p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right) p\left(\alpha_t \mid \theta^*, F_{t-1}\right) d\alpha_t \end{aligned} \tag{3.11}$$

As α_t is non-observable, the exact integral in Equation (3.11) using an analytical method is hard to obtain. However, with the help of Monte Carlo averaging $p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right)$ over a large sample of draws of $\alpha_t^1, ..., \alpha_t^M$ from $p\left(\alpha_t \mid \theta^*, F_{t-1}\right)$, we could have an approximation of $p\left(y_t \mid \theta^*, F_{t-1}\right)$ as the following:

$$p\left(y_t \mid \theta^*, F_{t-1}\right) \simeq \frac{1}{M} \sum_{g=1}^{M} p\left(y_t \mid \alpha_t^{(g)}, \theta^*, F_{t-1}\right) \quad (g \leq M)$$

Herein, a repeated two-stage filtering procedure is involved. In the next several paragraphs, we firstly introduce the filtering principle. Secondly, we introduce the Particle Filter, together with some sampling methods within the Particle Filter framework and the weakness of the Particle Filter. Finally, we introduce the APF, which is an evolutive form of the PF, along with a sampling method within the APF and the application algorithm of APF for the evaluation of GSTUR model's likelihood ordinates.

(1) Filtering Principle

The target sampling density $p\left(\alpha_t \mid \theta^*, F_{t-1}\right)$ can be derived from the previous density $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$ via the transition density $p\left(\alpha_t \mid \alpha_{t-1}\right)$.

$$\begin{aligned} p\left(\alpha_t \mid \theta^*, F_{t-1}\right) &= \int p\left(\alpha_t \mid \alpha_{t-1}, \theta^*, F_{t-1}\right) p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right) d\alpha_{t-1} \\ &= \int p\left(\alpha_t \mid \alpha_{t-1}, \theta^*, F_{t-1}\right) dF\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right) \end{aligned} \tag{3.12}$$

A numerical method can be applied to evaluate the moments of $p\left(\alpha_t \mid \theta^*, F_{t-1}\right)$ by taking samples from $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$.

Via the Bayes theorem,

$$p\left(\alpha_t \mid \theta^*, F_t\right) = \frac{p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right) p\left(\alpha_t \mid \theta^*, F_{t-1}\right)}{p\left(y_t \mid \theta^*, F_{t-1}\right)}$$

$$p\left(y_t \mid \theta^*, F_{t-1}\right) = \int p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right) dF\left(\alpha_t \mid \theta^*, F_{t-1}\right) \tag{3.13}$$

The current state $p\left(\alpha_t \mid \theta^*, F_t\right)$ can be propagated into the future as it shown in (3.12), where the previous state $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$ is propagated into the current state. This procedure is taken throughout the data for all the t.

(2) SIR-based Particle Filters

The Particle Filters (PF hereafter) are defined as

> "the class of simulation filters that recursively approximate the filtering random variable $\alpha_t \mid F_t$ by 'particles' $\alpha_t^1, \cdots, \alpha_t^M$, with discrete probability mass of $\pi_t^1, \cdots, \pi_t^M$. Hence a continuous variable is approximated by a discrete one with a random support. These discrete points are viewed as samples from $p\left(\alpha_t \mid F_t\right)$." – Pitt and Shephard (1999)

At the state of $t-1$, since the discrete support $\pi_{t-1}^1, \cdots, \pi_{t-1}^M$ of the particles $\alpha_{t-1}^1, \cdots, \alpha_{t-1}^M$ are treated as the true filtering density in the PF, we can produce an approximation to the density

$$\widehat{p}\left(\alpha_t \mid \theta^*, F_{t-1}\right) = \sum_{j=1}^{M} p\left(\alpha_t \mid \theta^*, \alpha_{t-1}^j\right) \pi_{t-1}^j \tag{3.14}$$

And then, step 1 in (3.13) can be approximated as

$$\widehat{p}\left(\alpha_t \mid \theta^*, F_t\right) \propto p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right) \sum_{j=1}^{M} p\left(\alpha_t \mid \theta^*, \alpha_{t-1}^j\right) \pi_{t-1}^j \tag{3.15}$$

If π_{t-1}^j are assumed to equal $1/M$, (3.15) will be just the same as combining a prior $p\left(\alpha_t \mid \theta^*, \alpha_{t-1}\right)$ and the likelihood $p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right)$ to obtain a posterior.

In the PF literature, important work includes the following: It is proposed that the algorithm PF be applied with nonlinear/non-Gaussian models in Gordon, Salmond and Smith (1993) in the field of signal processing. In the time series area, the PF algorithm was introduced in Kitagawa (1996) as a solution to nonlinear non-Gaussian higher dimensional state space models. Some statistical refinements and further investigations regarding PF can be found in Carpenter, Clifford and Fearnhead (1999) and Fearnhead (2002, 2004). In Pitt and Shephard (1999), different types of PF and their limitations were intensively reviewed and the novel APF was proposed.

We firstly introduce a Sampling/Importance Resampling (SIR) method. This SIR has been suggested in the PF framework in Gordon et al. (1993) and Kitagawa (1996). The idea of SIR is that if we can get a sample of $\left\{\alpha_{t-1}^1, \cdots, \alpha_{t-1}^R\right\}$ according to the

normal transition prior density $p(\alpha_{t-1} \mid \theta^*, \alpha_{t-2})$, when the new data y_{t-1} arrives and data information is updated from F_{t-2} to F_{t-1}[12], $\left\{\alpha_{t-1}^1, \cdots, \alpha_{t-1}^R\right\}$ is resampled M times by associating each of the draw with weights π^j, where

$$\begin{aligned} w_j &= p\left(y_{t-1} \mid \alpha_{t-1}^j, \theta^*, F_{t-2}\right), \\ \pi^j &= \frac{w_j}{\sum_{i=1}^R w_i}, j = 1, \cdots, R \end{aligned} \tag{3.16}$$

The resampled $\left\{\alpha_{t-1}^1, \cdots, \alpha_{t-1}^M\right\}$ with a sample size of M will then be an approximated sample from our target density $p(\alpha_{t-1} \mid \theta^*, F_{t-1})$. This SIR method requires $R \gg M$.

There are two main weakness in this SIR-based PF. The first weakness is the potential inefficiency of the SIR-based PF. When the conditional likelihood is highly peaked compared to the prior and π^j is very variable, the SIR-based PF would become very imprecise as many of the samples from $p(\alpha_{t-1} \mid \theta^*, \alpha_{t-2})$ will be in the region where the conditional likelihood $p\left(y_{t-1} \mid \alpha_{t-1}^j, \theta^*, F_{t-2}\right)$ is very small, and would therefore contribute little to the accuracy of the approximation of the posterior density. The second weakness is the poor tail approximation, which holds in general for all PFs if π^j are treated with equal weights. When there is an outlier in $p(\alpha_{t-1} \mid \theta^*, F_{t-2})$, the tail densities of $p(\alpha_{t-1} \mid \theta^*, F_{t-2})$ can only be poorly approximated by the discrete sample from $p(\alpha_{t-1} \mid \theta^*, \alpha_{t-2})$ in (3.14). Consequently, the approximation of the tails in (3.15) will be very poor.

These limitations of PF together with other sampling methods in the PF framework, such as the adapted SIR-based PF, the blind rejection sampling-based PF, and the blind MCMC method are all reviewed in Pitt and Shephard (1999).

(3) APF and its Application

APF is actually an evolved form of the Particle Filters. In the APF, the target is to sample from the joint density $p(\alpha_{t-1}, k \mid \theta^*, F_{t-1})$ rather than $p(\alpha_{t-1} \mid \theta^*, F_{t-1})$. The APF involves a two-stage reweighting and resampling by introducing the index, k, which is called an auxiliary variable. An importance function $g(\cdot)$ is also introduced, which plays an important role associating draws from the prior with the predictive likelihood.

The SIR-based APF is described as follows: we can make R proposals α_{t-1}^j, k^j : $j = 1, \cdots, R$

$$\alpha_{t-1}^j, k^j \sim g(\alpha_{t-1}, k \mid \theta^*, F_{t-1})$$

[12]This indicates that the states are updated one period at a time.

and the resampling weights are constructed as the follows:

$$\begin{aligned} w_j &= \frac{p\left(y_{t-1} \mid \alpha_{t-1}^j, \theta^*, F_{t-2}\right) p\left(\alpha_{t-1}^j \mid \theta^*, \alpha_{t-2}^{k^j}\right)}{g\left(\alpha_{t-1}^j, k^j \mid \theta^*, F_{t-1}\right)}, \\ \pi^j &= \frac{w_j}{\sum_{i=1}^R w_i}, j = 1, \cdots, R \end{aligned}$$

The importance function $g(\cdot)$ should be designed depending on the updated data y_{t-1} and also $\alpha_{t-2}^{k^j}$ to make the weights even. With $g(\cdot)$, α_{t-1}^j that associate with large predictive likelihood would receive higher weights. If we denote $E_{\alpha_{t-1}|\alpha_{t-2}}^k$ as the mean associated with the transition density of $p\left(\alpha_{t-1} \mid \alpha_{t-2}^k\right)$, according to Pitt and Shephard (1999), $g(\cdot)$ can be selected as the following approximation from Equation (3.17)

$$g\left(\alpha_{t-1}, k \mid \theta^*, F_{t-1}\right) \propto p\left(y_{t-1} \mid E_{\alpha_{t-1}|\alpha_{t-2}}^k, \theta^*, F_{t-2}\right) p\left(\alpha_{t-1} \mid \theta^*, \alpha_{t-2}^k\right) \pi^k, \qquad (3.17)$$

where $k = 1, \cdots, M$. This form of the approximating density $g(\cdot)$ is designed so that

$$\begin{aligned} g\left(k \mid F_{t-1}\right) &\propto \pi^k \int p\left(y_{t-1} \mid E_{\alpha_{t-1}|\alpha_{t-2}}^k, \theta^*, F_{t-2}\right) dF\left(\alpha_{t-1} \mid \theta^*, \alpha_{t-2}^k\right) \\ &\propto \pi^k p\left(y_{t-1} \mid E_{\alpha_{t-1}|\alpha_{t-2}}^k, \theta^*, F_{t-2}\right) \\ w_k &= p\left(y_{t-1} \mid E_{\alpha_{t-1}|\alpha_{t-2}}^k, \theta^*, F_{t-2}\right) \\ \pi^k &= \frac{w_k}{\sum_{i=1}^M w_i}, k = 1, \cdots, M \end{aligned}$$

where $F\left(\alpha_{t-1} \mid \theta^*, \alpha_{t-2}^k\right)$ is the corresponding c.d.f. Thus, we get samples of (α_{t-1}, k) from $g\left(\alpha_{t-1}, k \mid F_{t-1}\right)$ by simulating k with probability λ_k proportional to $g\left(k \mid F_{t-1}\right)$,

$$\lambda_k \propto g\left(k \mid F_{t-1}\right)$$

then associate the k with α_{t-1}^k and α_{t-2}^k. λ_k is the so called the first-stage weights. By the end of the first-stage weights, a discrete sample with a size of M is obtained from the target density $g\left(\alpha_{t-1}, k \mid F_{t-1}\right)$ with a random support.

In the second-stage, we sample the indices k for R times and index these samples as $\left(\alpha_{t-1}^j, k^j\right) : j = 1, \cdots, R$, where R is $5 - 10$ times larger than M recommended by Chib et al. (2006). By doing so, we have R samples from the joint density $g\left(\alpha_{t-1}, k \mid F_{t-1}\right)$. As we would like to get particle samples where predictive likelihoods are high, a reweighting can be performed by putting weights on the draw $\left(\alpha_{t-1}^j, k^j\right)$,

where the weights are proportional to the second-stage weights π^j

$$\begin{aligned} w_j &= \frac{p\left(y_{t-1} \mid \alpha_{t-1}^j\right)}{p\left(y_{t-1} \mid E_{\alpha_{t-1}|\alpha_{t-2}}^{k^j}\right)}, \\ \pi^j &= \frac{w_j}{\sum_{i=1}^R w_i}, \end{aligned}$$

The hope is that these second-stage weights π^j are much less variable than those using the original SIR-based PF in equation (3.16). Finally, we randomly get M samples from the R discrete draws $\left(\alpha_{t-1}^j, k^j\right)$.

The beauty of APF lies in the choice of $g(\cdot)$ and the auxiliary variable k. In the first-stage, the discrete samples are weighted to approximate the $g(\alpha_{t-1}, k \mid F_{t-1})$. In the second-stage, the weights are reweighted to associate with the likelihood $p\left(y_{t-1} \mid \alpha_{t-1}^j\right)$. According to the simulation results in Pitt and Shephard (1999), the APF is very flexible and reliable to approximate $p(\alpha_t \mid y_t)$ when the conditional likelihood $p(y_t \mid \alpha_t)$ is very sensitive to α_t. Algorithm 2 explains how the APF is applied with the GSTUR model for the likelihood ordinates estimation at θ^*.

Algorithm 2: Estimate the log Likelihood Ordinates at θ^* Using the Auxiliary Particle Filter

First, at time t, we call the lags of α_t as $\underline{\alpha}_t = (\alpha_{t-1}, \cdots, \alpha_{t-p})'$. The lags of α_2 denoted as $\underline{\alpha}_2^{(g)} : g = 1, \cdots, M$, which are the initial values that can be either set as a $M \times p$ zeros matrix or a sample of M draws $(\underline{\alpha}_2^{(1)}, \cdots, \underline{\alpha}_2^{(M)})'$ from the conditional prior $p(\underline{\alpha}_2|\theta^*)$. In my Matlab code, $\underline{\alpha}_2^{(g)} : g = 1, \cdots, M$ is set as a $M \times p$ zeros matrix and M is set as $3,000$.

1. t starts from $t = 2$.

   (a) For each $\underline{\alpha}_t^{(g)}$, $g = 1, \cdots, M$, sample a value $\alpha_t^{\triangle(g)}$ using the transition density:

   $$\alpha_t^{\triangle(g)} \sim f_N(\underline{\alpha}_t^{(g)}\phi, \sigma_\eta^{2*})$$

   Note that $\alpha_t^{\triangle(g)}$ is a sample from $p(\alpha_t|\theta^*, \underline{\alpha}_t)$.

   (b) An estimate of the likelihood ordinate $p(y_t|\theta^*, F_{t-1})$ is given by:

   $$\hat{p}(y_t|\theta^*, F_{t-1}) = \frac{1}{M}\sum_{g=1}^{M} p(y_t|\theta^*, \alpha_t^{\triangle(g)}, F_{t-1}) \tag{3.18}$$

2. For each $g = 1, \cdots, M$ define $\hat{\alpha}_t^g = E(\alpha_t^g | \underline{\alpha}_t^g) = \underline{\alpha}_t^g \phi$ and calculate:

$$\begin{aligned} w_g &= p(y_t | \hat{\alpha}_t^g, \theta^*, F_{t-1}) \\ \pi_k &= \frac{w_g}{\sum_{j=1}^M w_j}, k = 1, \cdots, M \end{aligned}$$

π_k is the first-stage weights. Get R draws $(k_1, \cdots, k_R)$ from the discrete distribution defined on the integers $(1, \cdots, M)$ with probabilities $\pi_1, \cdots, \pi_M$, where R is set as larger as $5 \times M$. This step is to get R samples from the importance function $g(\alpha_t, k \mid \theta^*, F_t)$ by simulating the index with probability π_k. Note that each value of k_r is used to indicate a value of $\underline{\alpha}_t^{(k_r)}$ (and of $\hat{\alpha}_t^{k_r}$) and k_r. Explicitly, step 2 is using the importance function $g(\cdot)$

$$\begin{aligned} g(\alpha_t, k \mid \theta^*, F_t) &\propto p\left(y_t \mid E^k_{\alpha_t | \underline{\alpha}_t}, \theta^*, F_{t-1}\right) p\left(\alpha_t \mid \theta^*, \underline{\alpha}_t^k\right) \pi^k \\ k &= 1, \cdots, M \end{aligned}$$

to get a sample of R draws.

3. For each $\underline{\alpha}_t^{(k_r)}$, $r = 1, \cdots, R$, draw a scalar $\alpha_t^{*(r)}$ using the transition density $p(\alpha_t | \theta^*, \underline{\alpha}_t)$ with

$$\alpha_t^{*(r)} \sim f_N(\underline{\alpha}_t^{k_r} \phi, \sigma_\eta^{2*}) \tag{3.19}$$

Note that $\left(\alpha_t^{*(r)} : r = 1, \cdots, R\right)$ and $(k_r : r = 1, \cdots, R)$ is a sample from the joint density $g(\alpha_t, k_r | \theta^*, F_t, \underline{\alpha}_t)$

4. Resample the $R \times 1$ vector $\left(\alpha_t^{*(r)} : r = 1, \cdots, R\right)'$ M times with probabilities π_r defined as:

$$\begin{aligned} w_r^* &= \frac{p(y_t | \theta^*, F_{t-1}, \alpha_t^{*r})}{p(y_t | \theta^*, F_{t-1}, \hat{\alpha}_t^{k_r})} \\ \pi_r &= \frac{w_r^*}{\sum_{r=1}^M w_r^*} \end{aligned} \tag{3.20}$$

π_r is the second-stage weights. Then, the resampled $M \times 1$ vector, which contains values $(\alpha_t^{(1)}, \cdots, \alpha_t^{(M)})'$ is (approximately) distributed as $p(\alpha_t | \theta^*, F_t, \underline{\alpha}_t)$. These second-stage weights are associated with the conditional likelihood by the importance function $g(\cdot)$. Hence, the APF is better than the SIR-based PF, in which α_t only relies on the transition density and $(\underline{\alpha}_t^{(1)}, \cdots, \underline{\alpha}_t^{(M)})'$. Stacking this sampled $(\alpha_t^{(1)}, \cdots, \alpha_t^{(M)})'$ on $\underline{\alpha}_t^{(g)} : g = 1, \cdots, M$. We have the updated lags of $\underline{\alpha}_{t+1}^g = \left(\alpha_t^{(g)}, \alpha_{t-1}^{(g)}, \cdots, \alpha_{t-(p-1)}^{(g)}\right) : g = 1, \cdots, M$

5. Fix $t = t + 1$ go to step 1(a) to get $\alpha_{t+1}^{\triangle(g)}$ from $\underline{\alpha}_{t+1}^{g}$ in step 4. Note that $\alpha_{t+1}^{\triangle(g)}$ are samples from $p(\alpha_{t+1}^{\triangle(g)}|\theta^*, F_t, \underline{\alpha}_{t+1}^{g})$ and following 1(b) to get $\hat{p}(y_{t+1}|\theta^*, F_t)$[13], until $t = n$.

6. Finally, the estimate of the log likelihood is

$$\log \widehat{p}\,(y \mid \theta^*) = \sum_{t=2}^{n} \log \widehat{p}\,(y_t \mid \theta^*, F_{t-1}) \tag{3.21}$$

It might be argued that if it is good enough to set $\underline{\alpha}_2^{(g)}$ as a $M \times p$ zeros matrix. Our explanation is as follows: if y_1 is deterministic, the particle filter starts with

$$\underline{\alpha}_2^{(g)} = \begin{pmatrix} \alpha_1^{(1)} & \cdots & \alpha_{2-p}^{(1)} \\ \alpha_1^{(2)} & \cdots & \alpha_2^{(2)} \\ \vdots & & \vdots \\ \alpha_1^{(3000)} & \cdots & \alpha_{2-p}^{(3000)} \end{pmatrix}$$

, which is a $3000 \times p$ zero matrix. Once $\alpha_2^{\triangle(g)} \sim f_N(\underline{\alpha}_2^{(g)}\phi^*, \sigma_\eta^{2*})$ is simulated from the transition density, $\alpha_2^{\triangle(g)}$ is a sample from $p(\alpha_2|\theta^*, \underline{\alpha}_2)$, where θ^* are the estimated values of θ from the posterior sampler. The likelihood ordinate at time $t = 2$ for y_2, $p(y_2|\theta^*, y_1)$ is then estimated as

$$\hat{p}(y_2|\theta^*, y_1) = \frac{1}{M} \sum_{g=1}^{M} p(y_2|\theta^*, \alpha_2^{\triangle(g)}, y_1) \tag{3.22}$$

Notice that $\alpha_2^{\triangle(g)}$ is a sample from $p(\alpha_2|\theta^*, \underline{\alpha}_2)$, hence $\frac{1}{M}\sum_{g=1}^{M} p(y_2|\theta^*, \alpha_2^{\triangle(g)}, y_1)$ can be a good estimate of $\hat{p}(y_2|\theta^*, y_1)$. M can be chosen as a larger value when estimate $\hat{p}(y_2|\theta^*, y_1)$. However, since we require R to be $5 - 10$ times larger than M recommended by Chib et al. (2006), a larger M will induce bigger computational costs.

To sum up, a particle filtering method could recursively deliver sequence of draws of $p\left(\alpha_2^{(g)} \mid \theta^*, F_2, \underline{\alpha}_2\right)$,..., $p\left(\alpha_t^{(g)} \mid \theta^*, F_t, \underline{\alpha}_t\right)$, ..., $p\left(\alpha_n^{(g)} \mid \theta^*, F_n, \underline{\alpha}_n\right)$. Note that there are other less efficient methods to estimate the likelihood ordinate $p(y \mid \theta^*)$. For example,

[13] An estimate of $p(y_{t+1}|\theta^*, F_t)$ is given by:

$$\hat{p}(y_{t+1}|\theta^*, F_t) = \frac{1}{M} \sum_{g=1}^{M} p(y_{t+1}|\theta^*, F_t, \alpha_{t+1}^{\triangle(g)})$$

one could get sample draws of $\alpha_2^{(g)}, \cdots, \alpha_n^{(g)} : g = 1, \cdots, M$ from the conditional prior $p(\alpha_2, ..., \alpha_n \mid \theta^*)$ and simply average the conditional likelihood $p(y \mid \theta, \alpha_2, ..., \alpha_n)$ over these draws. Even though this method would also deliver an estimate of the likelihood $p(y \mid \theta^*)$, it is less efficient than using the two-stage reweighting and resampling APF. Because this method shares the same disadvantages as the SIR-based PF that most of the draws from the prior may contribute little to the estimate of the likelihood when the likelihood is sensitive to the prior. The advantages of using an APF is that in the APF, a sample of $\alpha_t^{(g)} : g = 1, \cdots, M$ that receives high conditional predictive likelihood receives more weights and the sample of $\alpha_t^{(g)}$ is updated along with t and the y_t so that the estimate of the log likelihood ordinate at θ^* will be more accurate.

With the 'marginal likelihood identity' from (3.9), estimated posterior ordinates using a 'reduce Gibbs runs' from (3.10) and the estimated log likelihood ordinates from an APF, we are able to estimate the log marginal likelihood of any model of interest, such as GSTUR model with different specifications. We also would like to compare the simplest linear RW model with the complicated nonlinear GSTUR model. The marginal likelihood of the RW model is computational simpler and the integral can be evaluated analytically[14]. As in this case:

$$p_{RW}(y) = \int p(y|\sigma_\varepsilon^2) p(\sigma_\varepsilon^2) d\sigma_\varepsilon^2 \tag{3.23}$$

In the RW model, σ_ε^2 is the only parameter that has to be estimated. If $h_\varepsilon = \sigma_\varepsilon^{-2}$ and the prior chosen for h_ε follows a Gamma distribution $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, the marginal likelihood of the RW model would heavily depend on the values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$. From the model comparison aspect, it is sensible to choose the same prior for the common parameters amongst the competing models. The values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ will be chosen as the same as those for σ_ε^{-2} in the GSTUR model (see Table 3.1).

With the available marginal likelihoods for the model of interest, the Bayes factors for competing models can be evaluated.

3.3 Evaluations Using Artificial Data

In this section, we evaluated the efficiency of the developed GSTUR model's posterior simulator. A series of artificial data with a sample size of 120 is simulated from the

[14]Please refer to Appendix A for the derivation of the analytical integration with the RW model.

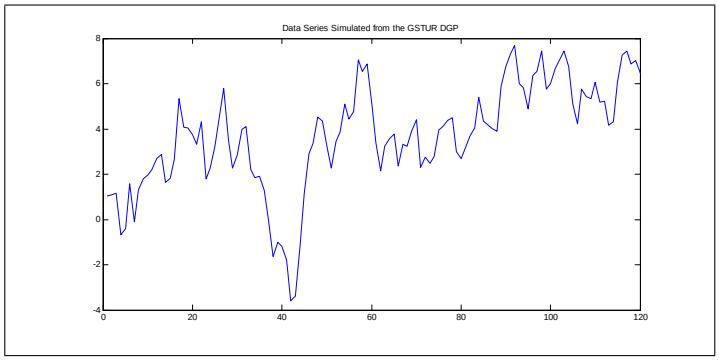


Figure 3-6: Simulated Data from GSTUR DGP

GSTUR DGP, specified as the following

$$\nu_t = y_t - \gamma - \delta t$$

where

$$\nu_t = \exp(\alpha_t)\nu_{t-1} + \sum_{i=1}^{l} \lambda_i \triangle \nu_{t-i} + \varepsilon_t \tag{3.24}$$

$$\alpha_t = \phi_0 + \phi_1 \alpha_{t-1} + \eta_t \tag{3.25}$$

where ε_t is $i.i.d.N(0, \sigma_\varepsilon^2)$ and η_t is $i.i.d.N\left(0, \sigma_\eta^2\right)$. Figure (3-6) plots the simulated series

3.3.1 Estimation

The 'true values' and estimated values are provided in Table (3.2). A GSTUR model with $p = 1$ and $l = 2$ is estimated by running the Gibbs Sampler for $25,000$ replications with the initial $5,000$ replications discarded. The estimated results for posterior means, standard deviations, CD diagnostics of the MCMC convergence and high density of the posterior interval are provided in Table (3.2). The 'true' values of the parameters are provided as a benchmark to evaluate the efficiency of the estimates.

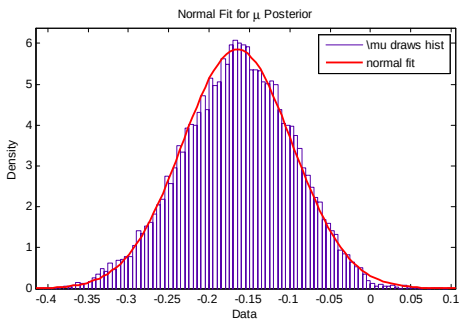
From Table (3.2), the estimated results for all parameters are close to the 'true' values and are within 99% High Posterior Density Intervals (HPDI). The Gibbs sampler converged for all the parameters according to the CD values. Therefore, it is not necessary to increase the iteration numbers as the convergence is achieved after $25,000$ replications. Fig (3.a-3.h) plots the posterior distributions of all the parameters of

Table 3.2: GSTUR Estimates Using Simulated Data

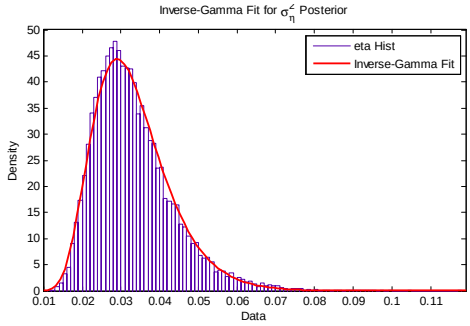
parameters	*trueV*	*Est.mean*	*s.t.d*	*CD*	*HPDI*.99	
μ_α	-0.2106	-0.165	0.0681	0.317	-0.3292	-0.0181
σ^2_η	0.0373	0.0332	0.0103	0.5601	0.0169	0.0662
σ^2_ε	0.919	1.1682	0.1645	0.2124	0.8457	1.6113
ϕ_1	-0.2276	0.1152	0.1326	-0.1481	-0.194	0.4249
γ	1	0.9731	1.5336	-0.5188	-3.0144	4.8222
δ	0.05	0.0446	0.0198	0.3196	-0.005	0.0946
λ_1	0.1619	0.1328	0.0983	-1.1332	-0.0971	0.3635
λ_2	-0.0165	-0.0214	0.1001	-0.0686	-0.2575	0.2126

parameters	NSE_{no}	$\text{NSE}_{.04}$	$\text{NSE}_{.08}$	$\text{NSE}_{.15}$	RNE_{no}	$\text{RNE}_{.04}$	$\text{RNE}_{.08}$	$\text{RNE}_{.15}$
μ_α	0.002	0.0284	0.0334	0.0326	1	0.0049	0.0035	0.0037
σ^2_η	0.0002	0.0007	0.0007	0.0007	1	0.0474	0.0484	0.0595
σ^2_ε	0.0047	0.0077	0.0075	0.0069	1	0.3761	0.3907	0.4611
ϕ_1	0.0013	0.0012	0.0011	0.0011	1	1.1066	1.2577	1.2777
γ	0.0167	0.0153	0.0132	0.0122	1	1.1949	1.5987	1.8747
δ	0.0002	0.0002	0.0002	0.0002	1	1	1.5285	1.6782
λ_1	0.0017	0.0089	0.0105	0.0103	1	0.0369	0.0268	0.0276
λ_2	0.0018	0.0106	0.0124	0.0121	1	0.0276	0.0205	0.0214

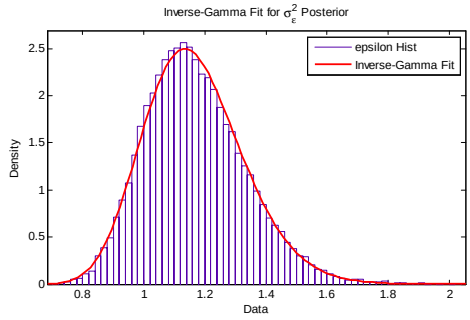
interest.



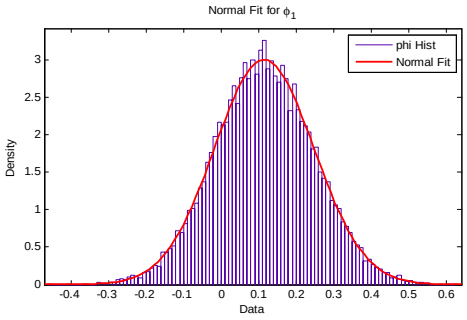
3.a



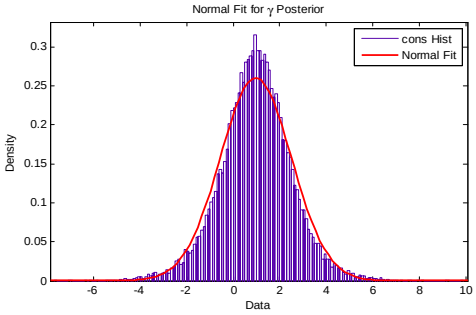
3.b



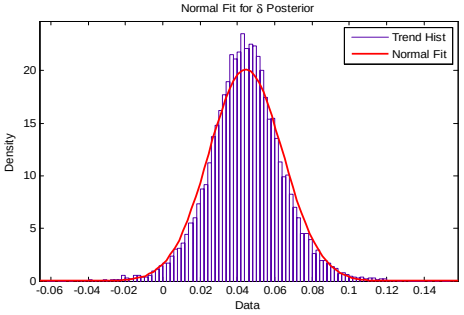
3.c



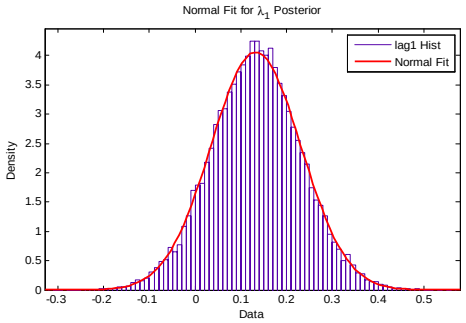
3.d



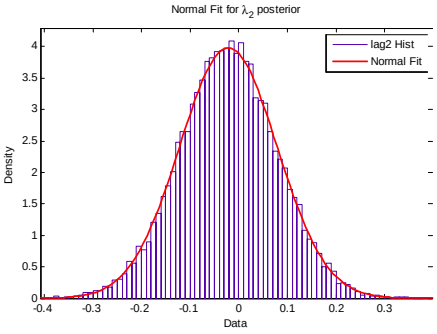
3.e



3.f



3.g



3.h

According to the experimental results under the controlled settings, the MCMC Gibbs sampler is able to provide efficient estimates for most of the parameters of interest. From Fig (3.a-3.h), normal posteriors can be well fitted with Normal distributions and the simulated posterior distributions of σ_η^2 and σ_ε^2 can be well fitted with the Inverse-Gamma distributions.

Table 3.3: Over-Dispersed Starting Points in 5 Chains

	$chain1$	$chain2$	$chain3$	$chain4$	$chain5$
μ_α	$\ln 0.9$	$\ln 0.9 + 3 \cdot 0.1$	$\ln 0.9 - 3 \cdot 0.1$	$\ln 0.9 + 5 \cdot 0.1$	$\ln 0.9 - 3 \cdot 0.1$
ϕ_1	1	$1 + 3 \cdot 0.1^{1/2}$	$1 - 3 \cdot 0.1^{1/2}$	$1 + 5 \cdot 0.1^{1/2}$	$1 - 3 \cdot 0.1^{1/2}$
Ω	$\binom{0}{0}$	$3 \cdot 10^2 \times \binom{1}{1}$	$-3 \cdot 10^2 \times \binom{1}{1}$	$5 \cdot 10^2 \times \binom{1}{1}$	$-5 \cdot 10^2 \times \binom{1}{1}$
λ	$zeros(l)$	$3 \cdot 10^2 \times ones(l)$	$-3 \cdot 10^2 \times ones(l)$	$5 \cdot 10^2 \times ones(l)$	$-5 \cdot 10^2 \times ones(l)$

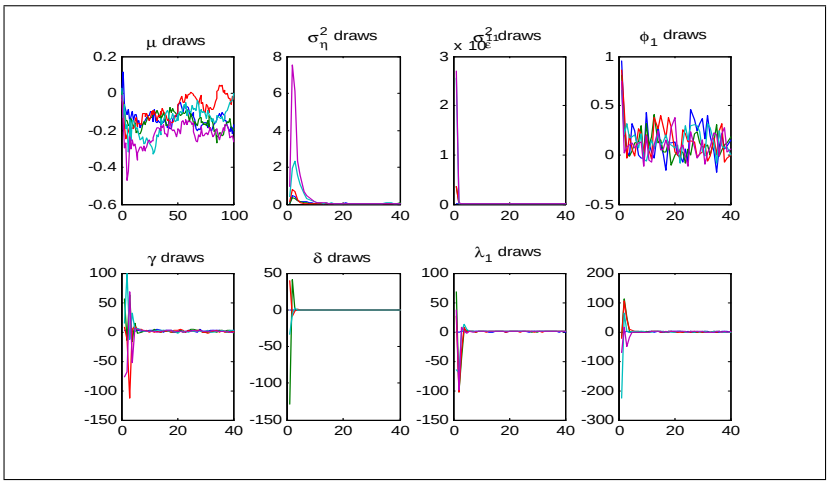


Figure 3-7: GSTUR 5 Parallel chains with Overdispersed Starting Points

3.3.2 MCMC with Dispersed Initial Values

From Table (3.2), the CD values show that the chain converges after the burn-in samples are discarded. To ensure the initial values do not contaminate the results, following the MCMC diagnostic procedure reviewed in Chapter 2.3.3, we run 5 parallel chains from different initial conditions and overlaying the sampled chains on a common graph for a visual inspection. The starting points are chosen as ± 5, ± 3 and 0 prior standard deviations from the prior means (see Table 3.3).

To illustrate the fast movements of the chains, we plot the first 100 MCMC samples for μ_α and first 40 MCMC samples for other parameters. From Fig (3-7), we may see the 5 parallel Markov chains converge for all the parameters within the first 40 iterations according to the visual inspection. Hence, the burn-in period with $5,000$ samples is big enough to avoid the contamination from the initial conditions.

Another way to visually inspect the MCMC efficiency is to plot the correlograms of the sample draws. An efficient MCMC requires small autocorrelations amongst the sample draws. Fig (3-8) plots the correlograms for chain 1.

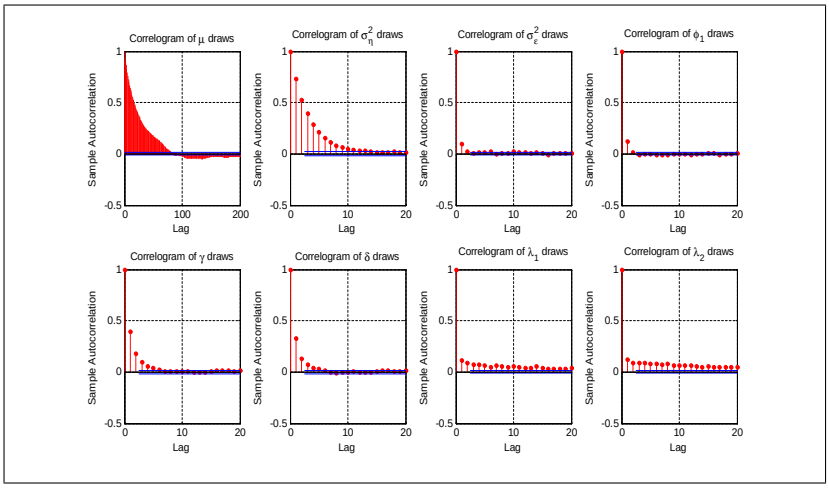


Figure 3-8: Correlograms: GSTUR parameter with 25,000 Iterations

From Fig (3-8), the correlogram (autocorrelation function) indicates small autocorrelations for most of the parameters. Only the correlogram of μ_α indicates autocorrelations at large (around 100) lag lengths. According to the derivations in Kim Shephard and Chib (1998), in the cases where the draws are uncorrelated, $10,000$ iterations is sufficient for an NSE to be less than 1% of the posterior standard deviation.

3.3.3 Prior Sensitivity Analysis

In this section, we employed much flatter priors for a simple prior sensitivity analysis. In Table (3.4), the prior variances (covariances) for μ_α, ϕ_1, Ω and λ are 100 times larger than those priors listed in Table (3.1).

From Table (3.4), the estimate of σ_ε^2 is sensitive to the prior choices. Whereas, the prior $\sigma_\varepsilon^2 \sim f_\Gamma^{-1}(1, 1/128)$ indicates a diffuse prior corresponding to a high degree of belief in σ_ε^2 being in the right tail, where $0 < \sigma_\varepsilon^2 < \infty$. As was mentioned in section 3.2.1, the values of the parameters in the prior $\sigma_\varepsilon^2 \sim f_\Gamma^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ cannot be chosen arbitrarily because the "priori" belief should at least to some extent, be a reflection of the real life data. The $\underline{\beta}_\varepsilon = 1/128$ in the prior $\sigma_\varepsilon^2 \sim f_\Gamma^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ is not a sensible one, since if we simulate data using a random sample σ_ε^2 from this prior, the hypothetical data will diverges from a reasonable range due to the large value of σ_ε^2. Hence, using different reasonable priors, as long as $\underline{\beta}_\varepsilon$ is not selected to be extremely small values, the estimates will not alter much.

Table 3.4: GSTUR Priors Sensitivity Analysis

priors	*trueV*	*Estmean*	*s.t.d*	*HPDI*.99	
$\mu_\alpha \sim f_N(\ln 0.9, 1)$	-0.2106	-0.3271	0.1915	-0.818	0.0283
$\sigma_\eta^2 \sim f_\Gamma^{-1}(1.5, 2.5)$	0.0373	0.044	0.0163	0.0207	0.0986
$\sigma_\varepsilon^2 \sim f_\Gamma^{-1}(1, 1/128)$	0.919	3.4371	0.4686	2.5254	4.7151
$\phi_1 \sim f_N(1, 10)$	-0.2276	0.1047	0.1285	-0.1954	0.4059
$\binom{\gamma}{\delta} \sim f_{MN}\left(\binom{0}{0}, \Sigma\right)$	$\begin{pmatrix} 1 \\ 0.05 \end{pmatrix}$	1.0021 0.0442	1.6088 0.0219	-3.4709 -0.0134	5.3345 0.1041
$\binom{\lambda_1}{\lambda_2} \sim f_{MN}\left(\binom{0}{0}, \Sigma\right)$	$\begin{pmatrix} 0.1619 \\ -0.0165 \end{pmatrix}$	0.1858 0.0353	0.1727 0.1753	-0.2205 -0.3781	0.5833 0.4466

para	NSE_{no}	$NSE_{.04}$	$NSE_{.08}$	$NSE_{.15}$	$RNE_{.04}$	$RNE_{.08}$	$RNE_{.15}$	CD
μ_α	0.0007	0.0036	0.0032	0.0033	0.0405	0.0522	0.0489	-1.3072
σ_η^2	0.0001	0.0003	0.0003	0.0002	0.1616	0.1891	0.2568	0.3926
σ_ε^2	0.0018	0.0023	0.0024	0.0022	0.6377	0.6131	0.7267	0.5403
ϕ_1	0.0015	0.0017	0.0017	0.0017	0.8031	0.7541	0.7569	-1.3711
γ	0.0176	0.026	0.0272	0.0275	0.4567	0.4179	0.407	1.0654
δ	0.0002	0.0003	0.0003	0.0003	0.6325	0.5938	0.5801	-0.3911
λ_1	0.0011	0.0017	0.0017	0.0018	0.4107	0.4106	0.3808	1.0841
λ_2	0.0011	0.0022	0.0019	0.0019	0.2597	0.3607	0.3529	1.2177

$$\Sigma = 10^6 eye\,(2)$$

3.3.4 Model Comparison via Bayes Factors

Under the controlled settings, using the Bayes Factors and the Chib method discussed in Section 2.3, we are able to focus on the existence of a deterministic time trend and the nonlinearity by investigating model probabilities for a group of competing models, which are specified as

1. GSTUR with constant and trend, where $\gamma \neq 0, \delta \neq 0$.
2. GSTUR with constant only, where $\gamma \neq 0$.
3. GSTUR with trend only, where $\delta \neq 0$.
4. GSTUR without constant and trend, where $\gamma = 0, \delta = 0$.
5. A pure RW model.

Table (3.5) presents the log marginal likelihood of the first four models and the Bayes Factors of GSTUR class models over the RW model. From Table (3.5), we see that the GSTUR model with $\delta \neq 0$ receives the highest marginal likelihood, which indicates the highest model probability amongst the four GSTUR class models of

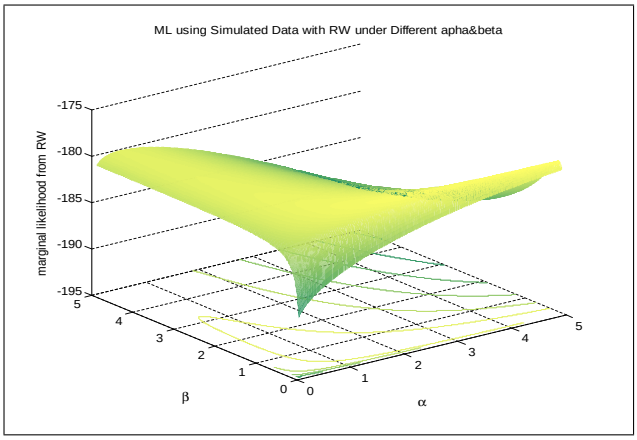


Figure 3-9: log Marginal Likelihood of RWwith Simulated Data from GSTUR DGP

interest. According to the log Bayes Factors (2.0887), the GSTUR model with $\delta \neq 0$ is supported when we compare it with the RW model.

Table 3.5: Model Comparison GSTUR VS RW Using a Common Prior with fixed $p = 1$ and $l = 2$

$GSTUR$	$\gamma \neq 0, \delta \neq 0$	$\gamma \neq 0$	$\delta \neq 0$	$\gamma = 0, \delta = 0$
$log\ ML$	-202.7103	-188.522	-179.051	-208.3532
$log\ BF_{GSTUR,RW}$	-21.5706	-17.3823	2.0887	-27.2135

Fig (3-9) plots the log marginal likelihood of a RW model with different selected priors of $\sigma_{\varepsilon}^{-2} \sim f_{\Gamma_{\varepsilon}}\left(\underline{\alpha}_{\varepsilon}, \underline{\beta}_{\varepsilon}\right)$, where the values of $\underline{\alpha}_{\varepsilon}$ and $\underline{\beta}_{\varepsilon}$ are ranged from 0.1 to 5 with a step of 0.01. The log marginal likelihood reaches the maximum point at -178.7182 when $\underline{\alpha}_{\varepsilon} = 5$ and $\underline{\beta}_{\varepsilon} = 0.17$. Under the same prior as that in the GSTUR model for the common parameter $\sigma_{\varepsilon}^{-2}$ by setting $\underline{\alpha}_{\varepsilon} = 1.1$ and $\underline{\beta}_{\varepsilon} = 0.2$, the log marginal likelihood using the RW will be -181.1397. Under the controlled setting using a common prior for $\sigma_{\varepsilon}^{-2}$, if we compare the GSTUR model, $M_{GSTUR:\delta\neq 0}$, with the RW model, M_{RW}, the log Bayes Factor, $\log BF_{GSTUR:\delta\neq 0,RW}$, is larger than 0, which indicates a support for the GSTUR model according to the Jeffreys classifications in Table (2.1). If we just consider two models M_i and M_j (which are RW and GSTUR with $\delta \neq 0$), the weight averaged model $\overline{M}$ can be expressed as

$$\overline{M} = M_i p(M_i) + M_j p(M_j) \tag{3.26}$$

and

$$p(M_j) = \frac{p(y \mid M_j)}{p(y \mid M_j) + p(y \mid M_i)} = \frac{1}{1 + Bij}, \tag{3.27}$$

where $p(M)$ indicates the model probabilities. We may conclude that, using the Bayes Factors, both a nonlinearity in the form of GSTUR and a deterministic time trend exist with a probability of 90% in the simulated data. This result is consistent with our DGP in (3.24 and 3.25).

3.4 Empirical Illustrations with a GSTUR Model

3.4.1 Empirical Results with Stock Price

To illustrate the model estimation and model comparison algorithms, a S&P500 annual data set, measured in logarithms, is chosen from the extended Nelson and Plosser's data set (from 1877 to 1988) for empirical applications. This data set has been previously tested for an exact Unit Root, deterministic time trend and changing persistence (see Nelson and Plosser 1982, Kwiatkowski et al. 1992, Gil-Alana and Robinson 1997). Whether this univariate time series is Trend Stationary (TS), Difference Stationary (DS) or neither, or if there is any change in the underlying processes cannot be concluded with certainty. This data set has also been applied by Jones and Marriott (1999) with the original Stochastic Unit Root model (the simplest form of the GSTUR in equation 3.1 and 3.2).

In this Chapter, not only the S&P 500 data is applied with the GSTUR model for estimations, but also the model probabilities are evaluated to shed light on the model uncertainties.

Estimation

To ensure that the effects of the starting values in the MCMC algorithms are insignificant, we take $25,000$ draws with the first $5,000$ discarded. The simulation efficiency is evaluated by the Numerical Standard Errors (NSE)[15], Convergence Diagnostic values (CD) and correlograms of the samples. Because the initial replications from the Gibbs Sampler are worn off, the MCMC convergence diagnostic results show that it is unnecessary to run the MCMC from dispersed starting values.

The correlogram (autocorrelation function) plots serial correlations of the draws from the posterior simulator. Fig (3-10 and 3-11) indicates that, for all the parameters

[15]NSE is reported taking the correlation up to lags of 15% of the size of the sample into account.

of interest, there is no significant autocorrelations at lag lengths larger than 15. Thus, the quick decaying autocorrelation indicates quick movements in the sampled draws. The Gibbs Sampler is efficient and the sampled draws converge fast to a consistent estimator. In the cases where the draws are uncorrelated, 10,000 iterations is sufficient for an NSE to be less than 1% of the posterior standard deviation. The iteration number we are taking, which is 25,000 draws with the first 5,000 discarded, is much larger than required to ensure the Gibbs Sampler converges.

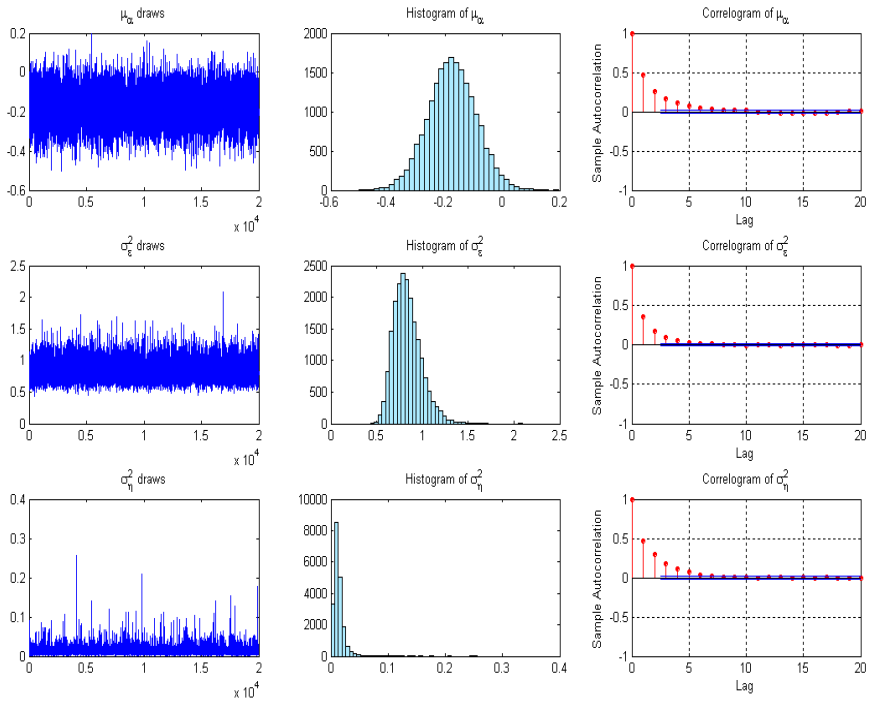


Figure 3-10: SP500 with GSTUR: Posterior Draws of μ_α, σ_ε^2 and σ_η^2

The summary statistics of Table (3.6) reports the estimated results and efficiency diagnostics of the Gibbs sampler. According to the CD value and NSE values, the Gibbs sampler converges for all the parameters of interest. The estimates from the Gibbs sampler could be sufficient for inferential purpose. The posterior distribution properties indicate the significances of the parameters of interest. A negative μ_α and small σ_η^2 indicate that the S&P500 series could be a good realization of a process with Stochastic Unit Roots.

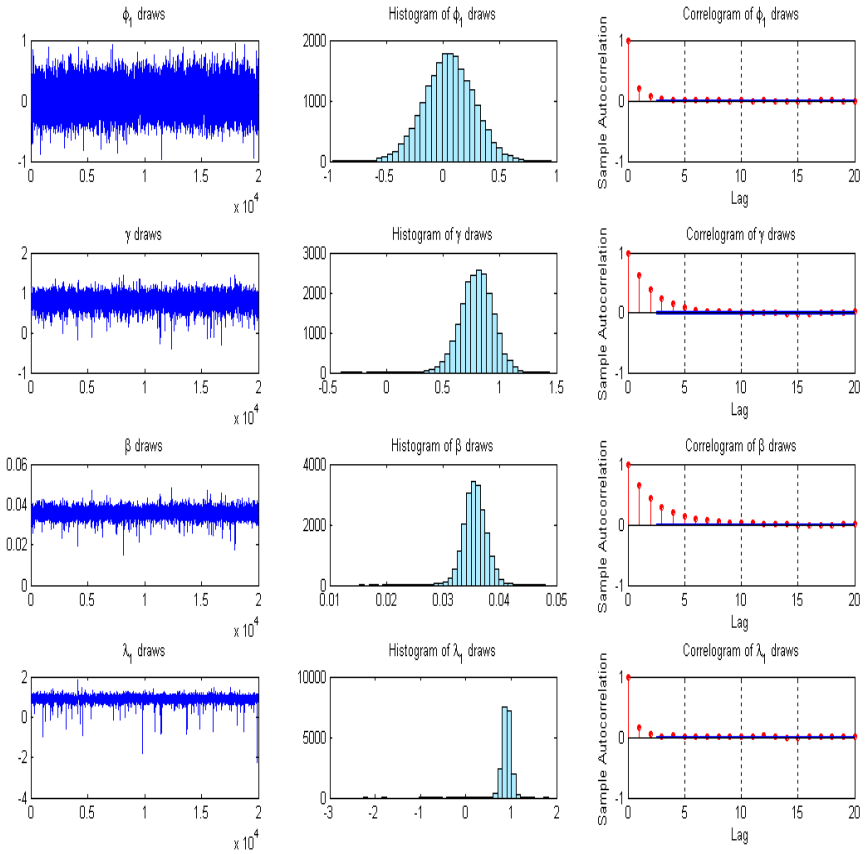


Figure 3-11: SP500 with GSTUR: Posterior Draws ϕ_1,γ, β, and λ_1

Table 3.6: Estimates: GSTUR with an Application to SP500

	Prior		*Posterior*					
	Mean	*StDev*	*Mean*	*StDev*	*CD*	*Median*	.95*HPDI*	
μ_α	$\ln 0.9$	0.1^2	-0.145	0.1036	0.0616	-0.137	-0.3259	0.0112
σ_η^2	-	-	0.0393	0.0131	-0.5059	0.0368	0.0229	0.0636
σ_ε^2	-	-	0.1167	0.016	-0.6674	0.1153	0.093	0.1449
ϕ_1	†	†	0.1094	0.1305	0.9928	0.1087	-0.1048	0.3255
γ	0	10^6	0.7657	0.3769	1.0057	0.7658	0.1415	1.3768
δ	0	10^6	0.0346	0.0054	-1.5527	0.0346	0.0256	0.0434
λ_1	0	10^6	0.2609	0.2128	-0.7117	0.2608	-0.0876	0.6115

parameters	NSE_{no}	$\text{NSE}_{.04}$	$\text{NSE}_{.08}$	$\text{NSE}_{.15}$	RNE_{no}	$\text{RNE}_{.04}$	$\text{RNE}_{.08}$	$\text{RNE}_{.15}$
μ_α	0.0011	0.01	0.0101	0.0112	1	0.0129	0.0126	0.0103
σ_η^2	0.0001	0.0005	0.0006	0.0005	1	0.0765	0.0683	0.0768
σ_ε^2	0.0002	0.0003	0.0003	0.0003	1	0.3753	0.3418	0.369
ϕ_1	0.0015	0.002	0.0023	0.0025	1	0.5321	0.3984	0.3537
γ	0.0042	0.0058	0.0056	0.0058	1	0.5189	0.5602	0.5248
δ	0.0001	0.0001	0.0001	0.0001	1	0.762	0.86	0.9842
λ_1	0.0024	0.0056	0.0059	0.0068	1	0.182	0.1599	0.121

† : $\phi_1 \sim f_N(0,1)\,1(\|z_j\| > 1)$ where $1(A)$ is the indicator function for the event A

− : see Table (3.1) for prior descriptions

Model Selection

Imposing $\delta = 0$ or $\gamma = \delta = 0$ significantly change the estimation results of the stochastic roots. Koop (1994) points that "imposing restrictions on the deterministic time trend is ruling out the possibility of a deterministic trend so that any trend behavior must manifest itself stochastically, biasing the tests in favour of stochastic nonstationarity". Considering over-parameterizing problems, it is also important to decide which parameters should be included for a good fitting model. In a Bayesian framework, the Bayes Factors is an effective tool in model selection procedures and can be applied to avoid the over-parameterizing problems (see Koop and Potter, 1999). The algorithms developed in Section 2, using the Chib method associated with the APF, are applied to calculate the model marginal likelihoods for an empirical application. Then, according to (2.1), the log Bayes Factors (logarithm of the marginal likelihood ratios) can be evaluated and model with the highest log marginal likelihood will be the most favoured model.

Table 3.7: log Marginal Likelihood with a GSTUR class of Models

	$\gamma \neq 0, \delta \neq 0$	$\gamma \neq 0$	$\delta \neq 0$	$\gamma = \delta = 0$
(a)	$p = 3$			
$l = 0$	-25.1406	-27.935	-4.7795	-66.5449
$l = 1$	-29.1619	-31.2879	-8.9852	-75.0414
$l = 2$	-36.9802	-39.3019	-16.77	-82.308
$l = 3$	-45.9399	-47.1458	-24.9254	-91.1269
(b)	$p = 2$			
$l = 0$	-23.9927	-26.4499	-2.7145	-62.7711
$l = 1$	-28.3142	-30.2792	-7.3366	-71.3285
$l = 2$	-35.5217	-37.6665	-14.7954	-79.0727
$l = 3$	-44.2955	-45.3416	-22.3381	-87.3335
(c)	$p = 1$			
$l = 0$	-21.6808	-23.4512	-0.2384*	-58.3054
$l = 1$	-26.0272	-27.51	-4.6843	-66.8312
$l = 2$	-33.4339	-34.7273	-11.6567	-74.8015
$l = 3$	-41.5651	-42.8262	-20.1654	-83.4481

Table (3.7) illustrates the estimated log marginal likelihood results. From Table (3.7), when the lag lengths are equal (same number of l) and under the same specifications of γ and δ, log marginal likelihoods of the GSTUR models do not alter a lot, given p varies from 1 to 3. Since p in the GSTUR model stands for the order of autoregressive in the hidden process, the transition equation (3.5) has little impact in determining the model fits. The model selection procedure, then, will highly depend on the specifications in the measurement equation (3.4).

To compare the highly parameterized nonlinear GSTUR model with a simple linear RW model, we calculate the marginal likelihood from the RW model using the S&P 500 data. Because the marginal likelihood from the RW model very much depends on the values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ in the prior $\sigma_\varepsilon^{-2} \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, which has been shown in Fig (3-9) under the controlled settings, we choose a range of values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ to calculate the marginal likelihood for a simple prior robust analysis. Both $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ start from 0.1 to 5 with a step of 0.01.

From Fig (3-12), the log marginal likelihood of a Random Walk model is maximized at 46.2606 with $\underline{\alpha}_\varepsilon = 5$ and $\underline{\beta}_\varepsilon = 5$. If the values are chosen as $\underline{\alpha}_\varepsilon = 1.1$ and $\underline{\beta}_\varepsilon = 0.2$,

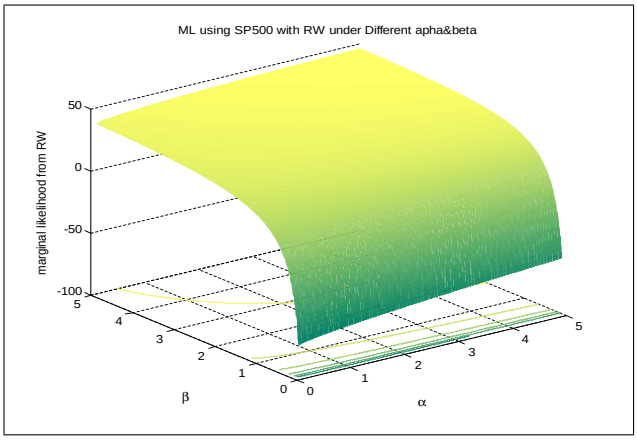


Figure 3-12: log Marginal Likelihood of RW with an Application of SP500

which are the same as those in the prior of $\sigma_{\varepsilon}^{-2} \sim f_{\Gamma_{\varepsilon}}\left(\underline{\alpha}_{\varepsilon}, \underline{\beta}_{\varepsilon}\right)$ in the GSTUR model, the log marginal likelihood is -34.413. Due to the fact that it is sensible to choose the same priors for the common parameters when we do the model comparison, under the same prior for $\sigma_{\varepsilon}^{-2}$, the Bayes Factors between the RW and the GSTUR model with a deterministic time trend ($\delta \neq 0$) can be calculated as:

$$BF_{RW:GSTUR} = \frac{p(M_{RW})}{p(M_{GSTUR})} = \frac{\exp(-34.413)}{\exp(-0.2384)} = 1.439\,3 \times 10^{-15}$$

According to Table (2.1), the coefficient nonlinearity in the form of stochastic unit roots is supported in the sample series. Taking the model uncertainties into account, if we just consider the RW model and the GSTUR (with $l = 1$, $\delta \neq 0$), we could weight average the probabilities of the roots being time-varying or being a constant according to equation (3.26 and 3.27). Therefore, we may conclude that the sample series has a 99% probability of being a stochastic unit root process.

To illustrate the possible changes of the persistence in the underlying process over the sample period, the estimated time-varying roots of α_t ($t = 1878 - 1988$) from the GSTUR model are plotted corresponding to the S&P 500 data. The estimates of the roots α_t vary under different specifications of the constant γ and the deterministic time trend δ. Fig (3-13) plots the stochastic roots estimated using a restricted GSTUR ($p = 1, l = 1$) model with no constant but a deterministic time trend, which is specified as $\delta \neq 0$. Fig (3-14) plots the stochastic roots estimated using a totally unrestricted GSTUR ($p = 1, l = 1$) model, which is specified with $\delta \neq 0$ and $\gamma \neq 0$.

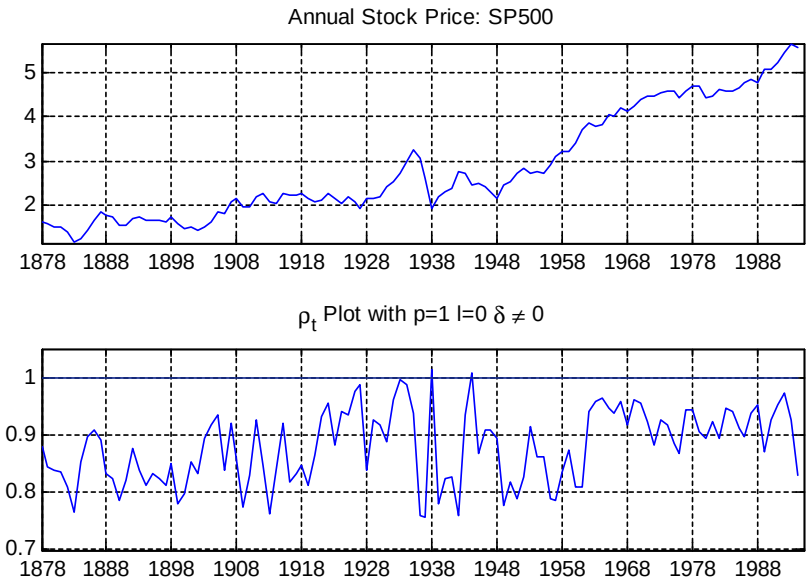


Figure 3-13: GSTUR with $\delta \neq 0$: Time-Varying Roots of SP500 1878-1988

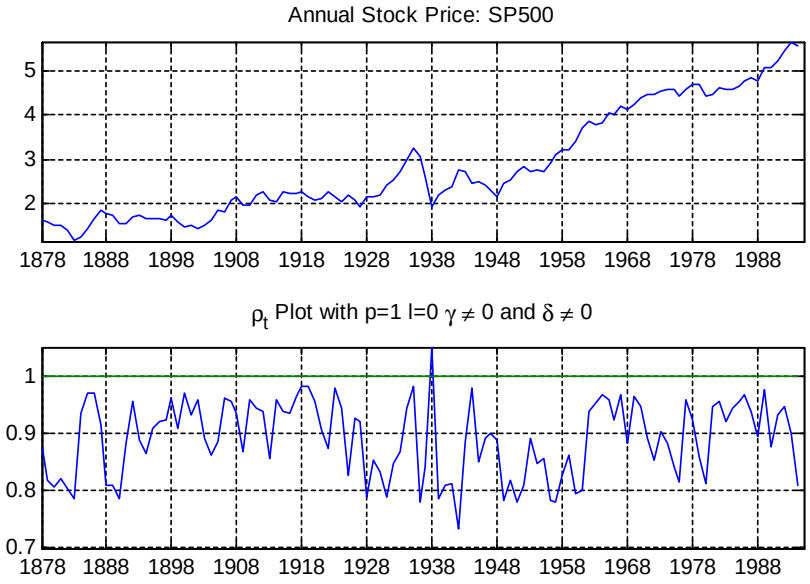


Figure 3-14: GSTUR with $\gamma \neq 0$ $\delta \neq 0$: Time-Varying Roots of SP500 1878-1988

From Fig (3-13 and 3-14), the roots go above unity occasionally. Explicitly speaking, the roots of the Stock prices series vary around $E[\exp(\mu_\alpha)]$ in the stationary region for most of the time, but go beyond 1 at certain time points and exhibit an explosive behaviour. At these time points, there might be changes in the persistence of the underlying process. The interval and the variances of the estimated stochastic roots α_t are small. These indicate that the STUR process may not be easily distinguished from a linear fixed unit root process. This may explain why Nelson and Plosser (1982), Kwiatkowski et al. (1992) all find evidences of a Unit Root in the S&P 500 series.

3.4.2 Empirical Results with Long-run Real Exchange Rate

Another empirical application for the GSTUR model is to analyze the real exchange rates. The data set applied in this chapter is the monthly U.K./U.S. real exchange rates from January 1885 to February 1995, with a sample size of $1,322$ over 111-year period. This data has been analyzed by Engel and Kim (1999) with a state space model. They applied an Augmented Dickey-Fuller (ADF) test and rejected the unit root in this long-span real exchange rate series at the 5 percent level.

In this section, a restricted GSTUR model ($\gamma = \delta = 0$, $p = 1$ and $l = 1$) is applied for analysis. The Gibbs Sampler uses $25,000$ iterations with the first $5,000$ discarded. The efficiency of the algorithm is shown according to the RNE values. The summary statistics of Table (3.8) reports the estimated results and diagnostics of the MCMC for this empirical application.

Figure (3-15) plots the U.K./U.S. real exchange rates, nominal exchange rates and estimated roots over 111 year span. According to the summaries in Engel and Kim (1999), we also provide the historical monetary events within the 111-year span:

1. Mid-1898. This period is congruent with the gold rush (Friedman and Schwartz, pp.135-28). During this period, nominal price levels in both the U.S. and the U.K. fluctuated wildly from month to month (while the nominal exchange rate remained fixed) as the money supplies of the two countries were jolted by increases in their stocks of gold.

2. Late 1902, there was a rapid inflation in the U.S. through September of 1902. This in part was driven by an expansionary monetary policy followed by the U.S. Treasury, which increased its deposits at major banks and persuaded the larger national banks to increase currency in circulation. The business cycle peaked in

Table 3.8: Estimates: GSTUR with U.K./U.S.Real Exchange Rates

	Prior		*Posterior*					
	Mean	*St.Dev*	*Mean*	*St.Dev*	*CD*	*Median*	95%*Posterior Band*	
μ_α	$\ln 0.9$	0.1^2	-0.0211	0.0510	0.0204	-0.0090	-0.1330	0.0276
σ_η^2	–	–	0.0002	0.0002	-0.1093	0.0001	0.0000	0.0006
σ_ε^2	–	–	0.0507	0.0020	-0.2337	0.0506	0.0475	0.0540
ϕ_1	†	†	0.4375	0.4855	-0.3223	0.5488	-0.5100	0.9812
λ_1	0	10^6	0.9461	0.0555	1.0901	0.9553	0.8416	1.0204

parameters	NSE_{no}	$\text{NSE}_{.04}$	$\text{NSE}_{.08}$	$\text{NSE}_{.15}$	RNE_{no}	$\text{RNE}_{.04}$	$\text{RNE}_{.08}$	$\text{RNE}_{.15}$
μ_α	0.0009	0.0012	0.0011	0.001	1	0.5891	0.7629	0.9316
σ_η^2	0	0	0	0	1	0.0556	0.0475	0.0499
σ_ε^2	0	0	0	0	1	1.0396	0.7532	0.5939
ϕ_1	0.0085	0.0164	0.0171	0.0158	1	0.2682	0.2487	0.291
λ_1	0.001	0.0027	0.0026	0.0021	1	0.1381	0.1511	0.2373

† : $\phi_1 \sim f_N(0,1)\,1(\|z_j\| > 1)$ where $1(A)$ is the indicator function for the event A
– : see Table (3.1) for description

September 1902, according to the NEBR. chronology, and prices fell from then until the end of the year(see Friedman and Schwartz, pp.149-52).

3. Mid-to-late-1910, As word War 1 began, the gold standard was abandoned. (See Grilli and Kaminsky, pp.193-94)

4. Mid-1919 to late-1920, the British ceased intervention in the foreign exchange market that had begun in December 1914, and the dollar began to float freely against the pound. This period of floating lasted until May 1925, as was noted above, but initially there was a large realignment of the exchange rate as the pound depreciated from $ 4.76 to less than $ 3.40 by February 1920. (See Grilli and Kaminsky, pp.194)

5. September 1931. Britain abandons the gold standard (Friedman and Schwartz, pp.380-84).

6. Mid-to late 1933. The United States ceases stabilizing the price of gold. The price of gold rises from $20.67 to $34 by January 1934 (Friedman and Schwartz, pp.462-91, and Grilli and Kaminsky, pp.194-95)

7. September 1939. Britain devalues the pound (See Grilli and Kaminsky, pp.195)

8. July 1946. Rapid U.S. inflation when price controls are removed (Friedman and Schwartz, pp.557-58)

9. September 1949, The pound is devalued (Grilli and Kaminsky, pp.195)

10. Late-1967, The pound is devalued.

11. Early-1985, Frankel (1994, pp.301-302) describes the period from June 1984 to February 1985 as a "speculative bubble" in the dollar. The price of the pound fell from about \$1.40 to less than \$1.10 during the period. The bubble burst in February 1985 and the exchange rate rose to near the \$1.40 by July of 1985.

12. Late-1992, The United Kingdom leaves the Exchange Rate Mechanism of the EMU. Between August and November 1992, the price of the pound falls from \$1.95 to \$1.52.

From Figure (3-15), the U.K./U.S. long run real exchange rate is highly persistent according to the estimated Stochastic Unit Roots. The range of the stochastic unit roots is narrow (from $0.98-1.015$), and the variance of the stochastic unit roots is small. For most of the time, the roots are below one, which indicates the series is stationary. At certain time points, the roots jump to or above one, which are marked with $\dotplus$. At these time points, the series exhibit nonstationary and/or explosive behaviors. We mark these time points, where the roots jump out of the unity, with $\dotplus$ in the plot of nominal exchange rates and real exchange rates. In the plot of nominal exchange rates, we find where the monetary events took place according to the list of historical record. The vertical bars in the plots indicate the time of significant monetary events. The stochastic roots, which are estimated to be above 1, are plotted together with the events together on the time line. We find that a change of the persistence in the series normally goes with an important monetary event sometime as shown in Fig (3-15). Since we might have omitted some other important events, the illustration of the correlation may not be significantly substantial.

3.5 Conclusions

For forecasting purposes, it is important to identify if a series is trend stationary or difference stationary, or neither. A coefficient nonlinear model, the GSTUR model is a flexible approach for modelling some macroeconomic time series'. Because in the GSTUR model, the roots are relaxed from a constant value to a process and the persistence in the series is allowed for changes with time. At this point, the GSTUR model could provide a better understanding and different implications about the source of macroeconomic fluctuations. The marginal likelihoods of the competing models are

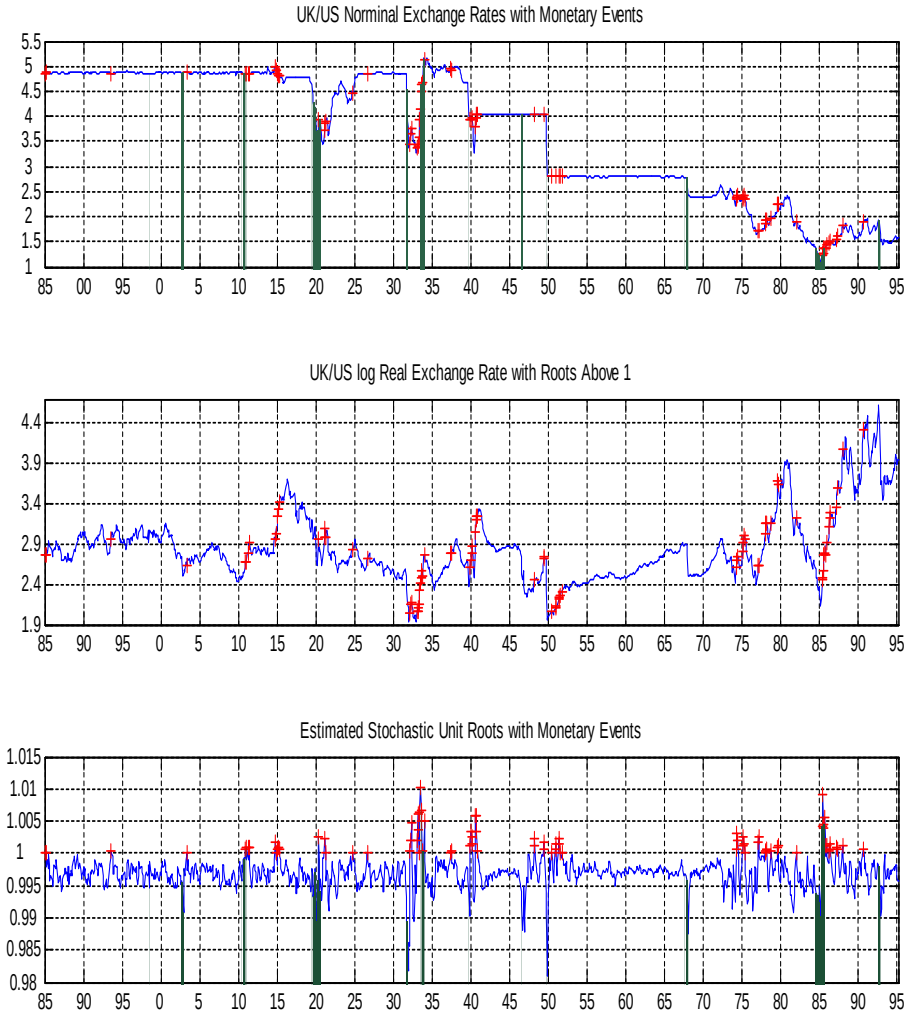


Figure 3-15: U.K./U.S. Long Run Exchange Rates and Estimated Stochastic Unit Roots

adequate to shed light on the model uncertainties and the existence of a deterministic time trend.

For Bayesian inferential purpose, with applications of the S&P 500 data sets and the U.K./U.S. long run real exchange rates, the Gibbs Sampler algorithm is efficient in providing consistent estimates for the highly parameterized dynamic GSTUR model. The 'efficient' Gibbs Sampler here, in particular, refers to a quick convergence in the MCMC algorithm, small serial correlations between the sample draws and fast movements in the sample draws. This MCMC efficiency is examined under controlled settings by the use of a simulated data series.

One of the main contributions of this chapter is to provide a complete set of tools for empirical practitioners, who wish to use the GSTUR model with empirical applications. This developed tool set contains a posterior simulator for estimation, a tool for MCMC diagnostics, model marginal likelihood evaluation, and estimations of the latent data α_t. Forecasting issues regarding the GSTUR will be discussed in more details in chapter 5. Considering the difficulties in the marginal likelihood evaluations for the high dimensional GSTUR model, where an analytical integral has a closed form cannot be obtained due to a $T \times 1$ dimension latent variables (α_t where $t = 2, \cdots, T$), this chapter implements the Chib method with an APF algorithm in this regards.

Another contribution of this chapter is a revisit of the dispute concerning the deterministic time trend in the S&P 500 series, which is part of the extended Nelson and Plosser's data set. An analysis using the GSTUR model suggests that the persistence has shifted within the sample. Comparing the computed log marginal likelihoods amongst the competing models, a GSTUR with a deterministic time trend model receives the highest marginal likelihood, which indicates a support of the deterministic time trend. Therefore, excluding the possibility of a deterministic trend may provide misleading inference. We propose that the underlying process of the S&P 500 series should be modelled with a more realistic approach, such as a combination of a deterministic time trend and a time varying persistence with roots varying stochastically.

Last, a simple analysis of the monthly U.K./U.S. long run real exchange rates over 111-year span suggests that a GSTUR model may help to resolve the PPP puzzle. The estimated time varying stochastic roots of the series suggest that important monetary events are connected to the shifts in the persistence of the real exchange rates.

Appendix 3.A Prior densities

According to the elicited priors, prior densities are expressed as follows

$$p\left(\mu_\alpha \mid \underline{\mu}_\alpha, \underline{V}_\alpha\right) = \frac{1}{\left(2\pi \underline{V}_\alpha\right)^{1/2}} \exp\left\{-\frac{1}{2\underline{V}_\alpha}(\mu_\alpha - \underline{\mu}_\alpha)^2\right\}$$

$$p(\lambda \mid \underline{\mu}_\lambda, \underline{V}_\lambda) = \frac{\left|\underline{V}_\lambda\right|^{-1/2}}{(2\pi)^{l/2}} \exp\left\{-\frac{1}{2}\left(\lambda - \underline{\mu}_\lambda\right)' \underline{V}_\lambda^{-1}\left(\lambda - \underline{\mu}_\lambda\right)\right\}$$

where $\lambda = (\lambda_1, \cdots, \lambda_l)'$.

$$p(\Omega \mid \underline{\mu}_\Omega, \underline{V}_\Omega) = \frac{\left|\underline{V}_\Omega\right|^{-1/2}}{(2\pi)^{2/2}} \exp\left\{-\frac{1}{2}\left(\Omega - \underline{\mu}_\Omega\right)' \underline{V}_\Omega^{-1}\left(\Omega - \underline{\mu}_\Omega\right)\right\}$$

$$p(\phi_i \mid \underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}) = \frac{1}{\underline{C}_\phi\left(2\pi \underline{V}_{\phi_i}\right)^{1/2}} \exp\left\{-\frac{1}{2\underline{V}_{\phi_i}}\left(\phi_i - \underline{\mu}_{\phi_i}\right)^2\right\}$$

where $\phi_i : i = 1, \cdots, p$ jointly satisfy the stationary condition and the integrating constant of prior $p(\phi)$ over the restricted region is $\underline{C}_\phi$.

$$p\left(h_\varepsilon \mid \underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) = \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right)$$

$$p\left(h_\eta \mid \underline{\alpha}_\eta, \underline{\beta}_\eta\right) = \frac{1}{\underline{\beta}_\eta^{\underline{\alpha}_\eta}\Gamma\left(\underline{\alpha}_\eta\right)} h_\eta^{\underline{\alpha}_\eta - 1} \exp\left(-\frac{h_\eta}{\underline{\beta}_\eta}\right)$$

For the values of $\left(\underline{\mu}_\alpha, \underline{V}_\alpha, \underline{\mu}_\lambda, \underline{V}_\lambda, \underline{\mu}_\Omega, \underline{V}_\Omega, \underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}, \underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon, \underline{\alpha}_\eta, \underline{\beta}_\eta\right)'$, refer to Table (3.1).

Appendix 3.B Posterior conditionals

The data series is denoted as $y = (y_{1-l}, \cdots, y_1, \cdots, y_n)'$ with a sample size of $T = n+l$, where the first $l+1$ values $(y_{1-l}, \cdots, y_1)'$ are treated as the starting values and the initial values of α is denoted as $\alpha_{initial} = (\alpha_{2-p}, ..., \alpha_1)'$, in which all elements equal to 0. We also denote $\alpha_{-t} = (\alpha_{2-p}, ..., \alpha_{t-1})'$.The joint density of the α is given by the sequence of conditionals as

$$p\left(\alpha|\theta\right) = \prod_{t=2}^{n} p\left(\alpha_t|\alpha_{-t}, \theta\right)$$

Since

$$\mu_\alpha = \frac{\phi_0}{1 - \sum\limits_{i=1}^{p} \phi_i}$$

$$\begin{aligned}\eta_t &= \alpha_t - \left[\phi_0 + \sum_{i=1}^{p} \phi_i \alpha_{t-i}\right] \\ &= \alpha_t - \left[\mu_\alpha \cdot \left(1 - \sum_{i=1}^{p} \phi_i\right) + \sum_{i=1}^{p} \phi_i \alpha_{t-i}\right] \\ &= (\alpha_t - \mu_\alpha) - \sum_{i=1}^{p} \phi_i \left(\alpha_{t-i} - \mu_\alpha\right)\end{aligned}$$

As we assume $\eta_t \sim f_N\left(0, \sigma_\eta^2\right)$, then we have

$$p\left(\alpha_t \mid \alpha_{-t}, \theta\right) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left\{-\frac{1}{2\sigma_\eta^2}\left[(\alpha_t - \mu_\alpha) - \sum_{i=1}^{p} \phi_i \left(\alpha_{t-i} - \mu_\alpha\right)\right]^2\right\}$$

and

$$\begin{aligned}p\left(\alpha \mid \theta\right) &= \prod_{t=2}^{n} p\left(\alpha_t | \alpha_{-t}, \theta\right) \\ &= \frac{1}{\left(2\pi\sigma_\eta^2\right)^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2\sigma_\eta^2} \sum_{t=2}^{n}\left[(\alpha_t - \mu_\alpha) - \sum_{i=1}^{p} \phi_i \left(\alpha_{t-i} - \mu_\alpha\right)\right]^2\right\} \quad (3.28)\end{aligned}$$

Given the likelihood function of $p\left(y|\alpha, \theta\right)$

$$p\left(y|\alpha, \theta\right) = \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{\frac{n-1}{2}}} \exp\left\{\begin{array}{c} -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^{n} \\ \left(\begin{array}{c} y_t - \exp(\alpha_t) y_{t-1} - \sum_{i=1}^{l} \lambda_i \triangle y_{t-i} - \\ \gamma\left[1 - \exp(\alpha_t)\right] - \\ \delta\left[t - \exp(\alpha_t) t + \exp(\alpha_t) - \sum_{i=1}^{l} \lambda_i\right] \end{array}\right)^2 \end{array}\right\} \quad (3.29)$$

The posterior is proportional to the product of prior and likelihood. With the specified priors $p\left(\theta\right) = p\left(\Omega, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2\right)$ and via the Bayes Theorem, the joint posterior density for (θ, α) is then

$$
\begin{aligned}
p(\theta,\alpha \mid y) &\propto \prod_{t=2}^{n} p(y_t \mid \alpha,\theta,F_{t-1})\, p(\alpha,\theta) \\
&= p(\alpha|\theta)\, p(\theta) \prod_{t=2}^{n} p(y_t \mid \alpha,\theta,F_{t-1}) \\
&\cdot \frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{\frac{n-1}{2}}} \exp\left\{ -\frac{1}{2\sigma_{\varepsilon}^{2}} \sum_{t=2}^{n} \left(\begin{array}{c} y_t - \exp(\alpha_t) y_{t-1} - \sum_{i=1}^{l} \lambda_i \triangle y_{t-i} - \\ \gamma\left[1-\exp(\alpha_t)\right] - \\ \delta\left[t - \exp(\alpha_t) t + \exp(\alpha_t) - \sum_{i=1}^{l} \lambda_i\right] \end{array} \right)^{2} \right\} \\
&\cdot \frac{1}{\left(2\pi\sigma_{\eta}^{2}\right)^{\frac{n-1}{2}}} \exp\left\{ -\frac{1}{2\sigma_{\eta}^{2}} \sum_{t=2}^{n} \left[(\alpha_t - \mu_\alpha) - \sum_{i=1}^{p} \phi_i \left(\alpha_{t-i} - \mu_\alpha\right) \right]^{2} \right\} \qquad (3.30)
\end{aligned}
$$

Then, we are able to develop the full conditional densities for the parameters of interest.

1. If we denote $h_\varepsilon = \frac{1}{\sigma_\varepsilon^2}$, and $\Omega := (\gamma, \delta)$, the posterior conditionals $\Omega \sim f_N\left(\overline{\mu}_\Omega, \overline{V}_\Omega\right)$ where

$$
\begin{aligned}
\overline{V}_\Omega &= \left(\underline{V}_\Omega^{-1} h_\varepsilon^{-1} + X'X\right)^{-1} h_\varepsilon^{-1} \\
&= \left(\underline{V}_\Omega^{-1} + X'X h_\varepsilon\right)^{-1} \\
\overline{\mu}_\Omega &= \overline{V}_\Omega \left(\underline{V}_\Omega^{-1} h_\varepsilon^{-1} \underline{\mu}_\Omega + X'X \widehat{\mu}_\Omega\right) \\
\widehat{\mu}_\Omega &= \left(X'X\right)^{-1} X'Y
\end{aligned}
$$

From (3.3), (3.5) and (3.4), the regression of Ω are with the following equation

$$
\begin{aligned}
& y_t - \exp(\alpha_t) y_{t-1} - \sum_{i=1}^{l} \lambda_i \triangle y_{t-i} \\
= \;& \gamma\left[1-\exp(\alpha_t)\right] + \delta\left[t - \exp(\alpha_t) t + \exp(\alpha_t) - \sum_{i=1}^{l} \lambda_i\right] + \varepsilon_t \qquad (3.31)
\end{aligned}
$$

If we take left hand side of (3.31) as Y

$$
Y = y_t - \exp(\alpha_t) y_{t-1} - \sum_{i=1}^{l} \lambda_i \triangle y_{t-i}
$$

and the right hand side of (3.31) as X

$$X = \left[1 - \exp(\alpha_t), t - \exp(\alpha_t)t + \exp(\alpha_t) - \sum_{i=1}^{l}\lambda_i\right]$$

where $t = 2, \cdots, n$, Regression Equation (3.31) became a simple linear regression equation

$$Y = X \cdot \Omega + \varepsilon_t$$

The posterior of Ω follows a normal distribution $f_N\left(\overline{\mu}_\Omega, \overline{V}_\Omega\right)$. See Koop (2003 pp.36-37) for details.

2. The posterior conditionals $\lambda \sim f_{MN}\left(\overline{\mu}_\lambda, \overline{V}_\lambda\right)$. The derivation of posterior conditionals is similar to that of Ω where

$$\begin{aligned}
\overline{V}_\lambda &= \left(\underline{V}_\lambda^{-1} h_\varepsilon^{-1} + X^{*\prime}X^*\right)^{-1} h_\varepsilon^{-1} \\
&= \left(\underline{V}_\lambda^{-1} + X^{*\prime}X^* h_\varepsilon\right)^{-1} \\
\overline{\mu}_\lambda &= \overline{V}_\lambda\left(\underline{V}_\lambda^{-1} h_\varepsilon^{-1}\underline{\mu}_\lambda + X^{*\prime}X^*\widehat{\mu}_\lambda\right) \\
\widehat{\mu}_\lambda &= \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^*
\end{aligned}$$

The regression of λ is with the following equation:

$$(y_t - \gamma - \delta t) - \exp(\alpha_t)\left[y_{t-1} - \gamma - \delta t + \delta\right] = \sum_{i=1}^{l}\lambda_i\left(\triangle y_{t-i} - \delta\right) + \varepsilon_t \tag{3.32}$$

if we denote

$$Y^* = (y_t - \gamma - \delta t) - \exp(\alpha_t)\left[y_{t-1} - \gamma - \delta t + \delta\right]$$

$$X^* = \left(\triangle y_{t-1} - \delta, \cdots, \triangle y_{t-i} - \delta\right)$$

where $t = 2, \cdots, n$, Equation (3.32) became

$$Y^* = X^* \cdot \lambda + \varepsilon_t$$

Then, we have posterior of λ follows a multivariate normal distribution $f_{MN}\left(\overline{\mu}_\lambda, \overline{V}_\lambda\right)$, where $\overline{\mu}_\lambda$ and $\overline{V}_\lambda$ can be obtained straight forward.

3. The posterior conditionals of error precision $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right)$. According to the prior density of h_ε

$$
\begin{aligned}
p(h_\varepsilon \quad | \quad y) &\propto \frac{h_\varepsilon^{\frac{n-1}{2}}}{2\pi^{\frac{n-1}{2}}} \exp\left\{-\frac{h_\varepsilon}{2}\left(Y - X\cdot\Omega\right)'\left(Y - X\cdot\Omega\right)\right\} \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) \\
&\propto h_\varepsilon^{\frac{n-1}{2}+\underline{\alpha}_\varepsilon - 1} \exp\left\{-h_\varepsilon\left[\frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2}\left(Y - X\cdot\Omega\right)'\left(Y - X\cdot\Omega\right)\right]\right\} \\
&\propto h_\varepsilon^{\overline{\alpha}_\varepsilon - 1} \exp\left\{-\frac{h_\varepsilon}{\overline{\beta}_\varepsilon}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\overline{\alpha}_\varepsilon &= \underline{\alpha}_\varepsilon + \frac{n-1}{2} \\
\overline{\beta}_\varepsilon^{-1} &= \frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2}\left(Y - X\cdot\Omega\right)'\left(Y - X\cdot\Omega\right)
\end{aligned}
$$

4. The posterior conditionals of $\mu_\alpha \sim f_N\left(\overline{\mu}_\alpha, \overline{V}_{\mu_\alpha}\right)$

To calculate the posterior of μ_α, collect all the parameters together

$$
\begin{aligned}
& p\left(\mu_\alpha \mid \alpha, \phi, h_\eta\right) \\
\propto\ & \exp\left\{-\frac{h_\eta}{2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2 - \frac{\left(\mu_\alpha - \underline{\mu}_\alpha\right)^2}{2\cdot\underline{V}_{\mu_\alpha}}\right\} \\
\propto\ & \exp\left\{-\frac{h_\eta}{2}\left(\begin{array}{c} \mu_\alpha^2\left[\sum_{t=2}^{n}\left(1 - \sum_{i=1}^{p}\phi_i\right)^2 + \frac{1}{h_\eta \underline{V}_{\mu_\alpha}}\right] \\ -2\mu_\alpha\left[\sum_{t=2}^{n}\left(\alpha_t - \sum_{i=1}^{p}\phi_i\alpha_{t-i}\right)\left(1 - \sum_{i=1}^{p}\phi_i\right) + \frac{\underline{\mu}_\alpha}{h_\eta \underline{V}_{\mu_\alpha}}\right] \end{array}\right)\right\}
\end{aligned}
$$

thus variance:

$$
\overline{V}_{\mu_\alpha} = \frac{1/h_\eta}{\sum_{t=2}^{n}\left(1 - \sum_{i=1}^{p}\phi_i\right)^2 + 1/h_\eta V_{\mu_\alpha}}
$$

From the joint posterior density function, it is easy to get the full conditional density of μ_α, which is with a mean of $\overline{\mu}_\alpha$

$$
\overline{\mu}_\alpha = \frac{\left(1 - \sum_{i=1}^{p}\phi_i\right)\sum_{t=2}^{n}\left(\alpha_t - \sum_{i=1}^{p}\phi_i\alpha_{t-i}\right) + \underline{\mu}_\alpha / h_\eta \underline{V}_{\mu_\alpha}}{\sum_{t=2}^{n}\left(1 - \sum_{i=1}^{p}\phi_i\right)^2 + 1/h_\eta V_{\mu_\alpha}}
$$

5. The posterior conditionals of ϕ_i

As the chosen prior for ϕ_i is $f_N\left(\underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}\right)$ and $\phi = \left(\phi_1, \cdots, \phi_p\right)'$ are jointly to meet the stationary condition, it is straight forward to derive the posterior conditionals for $\phi_i : i = 1, \cdots, p$ $\phi_i \sim f_N\left(\overline{\mu}_{\phi_i}, \overline{V}_{\phi_i}\right)$, with mean

$$\overline{\mu}_{\phi_i} = \frac{\sum_{t=2}^{n}\left(\alpha_{t-j} - \mu_\alpha\right)\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{k\neq j}^{p}\phi_k\left(\alpha_{t-k} - \mu_\alpha\right)\right] + \frac{\sigma_\eta^2 \underline{\mu}_{\phi_1}}{\underline{V}_{\phi_1}}}{\sigma_\eta^2/\underline{V}_{\phi_1} + \sum_{t=2}^{n}\left(\alpha_{t-j} - \mu_\alpha\right)^2}$$

and variance

$$\overline{V}_{\phi_i} = \frac{\sigma_\eta^2}{\sigma_\eta^2/\underline{V}_{\phi_i} + \sum_{t=2}^{n}\left(\alpha_{t-j} - \mu_\alpha\right)^2}$$

The derivation for ϕ_1 are as follows:

$$\begin{aligned}
p(\phi_1 \quad | \quad y) &\propto \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2 - \frac{1}{2}\sum_{i=1}^{p}\frac{\left(\phi_i - \underline{\mu}_{\phi_i}\right)^2}{\underline{V}_{\phi_i}}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{k=2}^{p}\phi_k\left(\alpha_{t-k} - \mu_\alpha\right) - \phi_1\left(\alpha_{t-1} - \mu_\alpha\right)\right]^2 - \frac{\left(\phi_1 - \underline{\mu}_{\phi_1}\right)^2}{2\underline{V}_{\phi_1}}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_\eta^2}\left[\begin{array}{c}\phi_1^2\frac{\sigma_\eta^2}{\underline{V}_{\phi_1}} + \phi_1^2\sum_{t=2}^{n}\left(\alpha_{t-1} - \mu_\alpha\right)^2 - 2\phi_1\frac{\sigma_\eta^2\underline{\mu}_{\phi_1}}{\underline{V}_{\phi_1}} \\ -2\phi_1\sum_{t=2}^{n}\left(\alpha_{t-1} - \mu_\alpha\right)\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{k=2}^{p}\phi_k\left(\alpha_{t-k} - \mu_\alpha\right)\right]\end{array}\right]\right\}
\end{aligned}$$

Then we can have the variance

$$\overline{V}_{\phi_1} = \frac{\sigma_\eta^2}{\sigma_\eta^2/\underline{V}_{\phi_1} + \sum_{t=2}^{n}\left(\alpha_{t-1} - \mu_\alpha\right)^2}$$

and the mean

$$\overline{\mu}_{\phi_1} = \frac{\sum_{t=2}^{n}\left(\alpha_{t-1} - \mu_\alpha\right)\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{k\neq 1}^{p}\phi_k\left(\alpha_{t-k} - \mu_\alpha\right)\right] + \frac{\sigma_\eta^2 \underline{\mu}_{\phi_1}}{\underline{V}_{\phi_1}}}{\sigma_\eta^2/\underline{V}_{\phi_1} + \sum_{t=2}^{n}\left(\alpha_{t-1} - \mu_\alpha\right)^2}$$

The derivations are all the same for $i = 1, \cdots, p$.

6. The posterior conditionals for h_η

$$
\begin{aligned}
p\left(h_\eta \mid \alpha, \mu_\alpha, \phi\right) &\propto \frac{1}{\underline{\beta}_\eta^{\underline{\alpha}_\eta}\Gamma\left(\underline{\alpha}_\eta\right)} h_\eta^{\underline{\alpha}_\eta - 1} \exp\left(-\frac{h_\eta}{\underline{\beta}_\eta}\right) \frac{h_\eta^{\frac{n-1}{2}}}{(2\pi)^{\frac{n-1}{2}}} \\
&\quad \exp\left\{-\frac{h_\eta}{2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2\right\} \\
&\propto h_\eta^{\underline{\alpha}_\eta + \frac{n-1}{2} - 1} \exp\left\{-h_\eta\left(\frac{1}{\underline{\beta}_\eta} + \frac{1}{2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2\right)\right\} \\
&\propto h_\eta^{\overline{\alpha}_\eta - 1} \exp\left\{-\frac{h_\eta}{\overline{\beta}_\eta}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\overline{\alpha}_\eta &= \underline{\alpha}_\eta + \frac{n-1}{2} - 1 \\
\overline{\beta}_\eta^{-1} &= \frac{1}{\underline{\beta}_\eta} + \frac{1}{2}\sum_{t=2}^{n}\left[\left(\alpha_t - \mu_\alpha\right) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2
\end{aligned}
$$

7. The remaining conditionals that are needed for the α_t are described as follows:

The conditional densities for α_t are nonstandard and given by the following expression

$$
\begin{aligned}
& p\left(\alpha_t \mid \Omega, \sigma_\varepsilon^2, \sigma_\eta^2, \lambda, \mu_\alpha, \phi, \alpha_{-t}, y\right) \\
\propto\ & \exp\left[-\frac{\nu_{t-1}^2}{2\sigma_\varepsilon^2}\left(e^{\alpha_t} - \frac{\nu_t - \sum_{i=1}^{l}\lambda_i \triangle \nu_{t-i}}{\nu_{t-1}}\right)^2 - \frac{\vartheta\left(\alpha_t\right)}{2\sigma_\eta^2}\left(\alpha_t - \frac{\tau\left(\alpha_t\right)}{\vartheta\left(\alpha_t\right)}\right)^2\right] \\
\text{where } \nu_t &= y_t - \gamma - \delta t \\
& \text{and } \vartheta\left(\phi\right) \text{ is a function of } \phi_i \\
& \tau\left(\alpha_t\right) \text{is a function of } \mu_\alpha, \phi_i, \alpha_{-t}, p \text{ and } t
\end{aligned}
$$

Derivations are as follows:

$$p\left(\alpha_1,\cdots,\alpha_n \mid \theta, y\right) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\underbrace{\left[\nu_t - \sum_{i=1}^{l}\lambda_i \triangle \nu_{t-i} - e^{\alpha_t}\nu_{t-1}\right]^2}_{A}\right\}$$
$$\cdot \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}\underbrace{\left[(\alpha_t - \mu_\alpha) - \sum_{i=1}^{p}\phi_i\left(\alpha_{t-i} - \mu_\alpha\right)\right]^2}_{B}\right\}$$

Next, we can rearrange it to make it looks nicer, rearrange into the form in Marriott's paper

$$A = -\frac{\nu_{t-1}^2}{2\sigma_\varepsilon^2}\left(e^{\alpha_t} - \frac{\nu_t - \sum_{i=1}^{l}\lambda_i \triangle \nu_{t-i}}{\nu_{t-1}}\right)^2$$

when $p = 1$

$$B = -\frac{1}{2\sigma_\eta^2}\sum_{t=2}^{n}\left[(\alpha_t - \mu_\alpha) - \phi_1\left(\alpha_{t-1} - \mu_\alpha\right)\right]^2$$

For t from 2 to $n-1$

$$\begin{aligned} B &= \left[(\alpha_t - \mu_\alpha) - \phi_1\left(\alpha_{t-1} - \mu_\alpha\right)\right]^2 + \left[(\alpha_{t+1} - \mu_\alpha) - \phi_1\left(\alpha_t - \mu_\alpha\right)\right]^2 + \cdots \\ &= \alpha_t^2\left(1 + \phi_1^2\right) - 2\alpha_t\left[\mu_\alpha\left(1 - \phi_1\right)^2 + \phi_1\left(\alpha_{t-1} + \alpha_{t+1}\right)\right] \\ &= \alpha_t^2\vartheta\left(\alpha_t\right) - 2\alpha_t \cdot \tau\left(\alpha_t\right) \end{aligned}$$

where

$$\vartheta\left(\alpha_t\right) = 1 + \phi_1^2$$

and

$$\tau\left(\alpha_t\right) = \mu_\alpha\left(1 - \phi_1\right)^2 + \phi_1\left(\alpha_{t-1} + \alpha_{t+1}\right)$$

For $t = n$

$$\begin{aligned} B &= \left[(\alpha_t - \mu_\alpha) - \phi_1\left(\alpha_{t-1} - \mu_\alpha\right)\right]^2 \\ &= \alpha_t^2\vartheta\left(\alpha_t\right) - 2\alpha_t \cdot \tau\left(\alpha_t\right) \end{aligned}$$

where

$$\vartheta\left(\alpha_t\right) = 1$$

and

$$\tau(\alpha_t) = \phi_1 \alpha_{t-1} + \mu_\alpha (1 - \phi_1)$$

When $p \geq 2$, the whole derivation procedure is the same and we summarize in the following table, which is also available in Jones and Marriott (1999)

Table 3.9: Appendix: Functions for Sampling apha (a)

values for $\vartheta(\phi)$ $p = 1$	
t	$\vartheta(\phi)$
$t \in [2, n-1]$	$1 + \phi_1^2$
$t = n$	1
values for $\tau(\mu_\alpha)$ $p = 1$	
t	$\tau(\mu_\alpha)$
$t \in [2, n-1]$	$\phi_1 (\alpha_{t-1} + \alpha_{t+1}) + \mu_\alpha (1 - \phi_1)^2$
$t = n$	$\phi_1 \alpha_{n-1} + \mu_\alpha (1 - \phi_1)$

Since part B can be approximated using a t-density, we used Independent Chain M-H algorithm to sample α_t. To generate draws with high acceptance probabilities, the selected candidate generating density should be with tails at least as fat as that of the posterior. I have chosen the degree of freedom as $\nu = 1$ in the t-density, which allows the t-density to have very fat tails.

Table 3.10: Appendix: Functions for Sampling apha(b)

values for $\vartheta(\phi)$ $p \geqslant 2$	
t	$\vartheta(\phi)$
$t \in [2, n-p]$	$1+\sum_{i=1}^{p}\phi_i$
$t \in (n-p, n-1]$	$1+\sum_{i=1}^{n-t}\phi_i$
$t = n$	1

values for $\tau(\mu_\alpha)$ $p \geqslant 2$	
t	$\tau(\mu_\alpha)$
$t \in [2, n-p]$	$\phi_p(\alpha_{t-p}+\alpha_{t+p})+\sum_{i=1}^{p-1}\left(\phi_i-\sum_{j=1}^{p-i}\phi_j\phi_{i+j}\right)(\alpha_{t-i}+\alpha_{t+i})+\mu_\alpha\left(1-\sum_{i=1}^{p}\phi_i\right)^2$
$t = n-1$	$\left[\sum_{k=1}^{p-1}\left(\phi_k-\phi_1\phi_{k+1}\right)\alpha_{t-k}\right]+\phi_p\alpha_{n-p-1}+\phi_1\alpha_n+\mu_\alpha(1-\phi_1)\left(1-\sum_{i=1}^{p}\phi_i\right)$
$t = n$	$\sum_{i=1}^{p}\phi_i\alpha_{N-i}+\mu_\alpha\left(1-\sum_{i=1}^{p}\phi_i\right)$

values for $\tau(\mu_\alpha)$ $p \geqslant 3$	
t	$\tau(\mu_\alpha)$
$t \in (n-p, n-2]$	$\phi_{n-t}\alpha_n+\phi_p\alpha_{t-p}+\sum_{k=1}^{p-1}\left(\phi_k-\sum_{m=1}^{\min(n-t,p-k)}\phi_m\phi_{k+m}\right)\alpha_{t-k}$ $+\sum_{k=1}^{n-t-1}\left(\phi_k-\sum_{m=1}^{n-t-k}\phi_m\phi_{k+m}\right)\alpha_{t+k}+\mu_\alpha\left(1-\sum_{i=1}^{p}\phi_i\right)\left(1-\sum_{k=1}^{n-t}\phi_k\right)$

Chapter 4

Evidence of Nonlinearity in the Inflation Rates with a Stationary Bilinear model

4.1 Introduction and Motivation

In this chapter, Bayesian methods are applied to a Stationary Bilinear (SB) model, in which the autoregressive (AR) coefficients are restricted within, and close to, the boundaries required for stationarity. The special feature of the SB model is that the first order AR coefficients are not only time variant but also are correlated with the random error terms in an economic context.

By adopting the Bayesian approach, a posterior simulator for model estimation, with a toolkit, which includes the model marginal likelihood evaluation and model forecasting, is developed. In the Bayesian context, a model selection between two non-nested models, which are the nonlinear SB model and the Random Walk (RW) model, is easily achievable via the Bayes factors.

Via experiments using the simulated data, we demonstrate that classical hypothesis testing and Bayesian model comparison may arrive at different conclusions. In the classical framework, a unit root in the series cannot be rejected using the conventional Augmented Dickey Fuller (ADF) test and the Phillips-Perron (PP) test. However, in the Bayesian context, compared with a RW model, the SB model based on the available data information receives a higher model probability according to the Bayes factors. A better fit of a Stationary Bilinear model leads us to question the feasibility of the conventional ADF test and the PP test with a nonlinear, parameter time-varying procedure.

For empirical illustrations, we applied the quarterly UK inflation rate from 1957 quarter one to 2007 quarter one. We found that the SB model could better represent the underlying process of the inflation series than a RW model.

The outline of this chapter is as follows: In this introduction and motivation section, we will elaborate the literature regarding inflation modelling. Also, we motivate the uses of the SB model and the Bayesian techniques via a survey of the literature. Section 2 gives the constructed proper priors and the forms of the corresponding posterior conditionals. Section 3 focuses on the methodology issues involved for analyzing the SB model, which include detailed discussions regarding the model estimation and the marginal likelihood evaluations. Section 4 evaluates the Markov chain Monte Carlo (MCMC) sampling efficiency with an application of a simulated data series in a controlled setting. In section 4, to evaluate the estimation efficiency with concern to the samples sizes, we also calculate and compare the Root Mean Square Errors (RMSE) of the estimates from the classical approach and the Bayesian approach. Section 5 uses the UK inflation rates data series for an empirical application. Finally, Section 6 concludes.

4.1.1 A Brief Review of Inflation Modelling

In this chapter, the UK headline inflation series is studied with a purely univariate statistical model, the Stationary Bilinear (SB) model. Based on the special features of the SB model, which is going to be reviewed in section 4.1.2, we pursue the research driven by the following questions: How persistent is the headline inflation rate? Do shocks in the economy have impacts on inflation rate persistence? Can we rely on the conventional ADF and PP tests results, to explain the persistence of the inflation time series?

As motivation, we address the importance of inflation modelling and some research puzzles in this area. Understanding inflation behaviour and its consequences are important subjects in modern macroeconomics because high rates of inflation reduce welfare (Lucas, 2000). Further more, high and persistent inflation undermines public confidence in the economy and in the management of economic policy (Bernake, 2007). In the long term, low inflation promotes growth, efficiency, and stability–which, all else being equal, support maximum sustainable employment (Bernake, 2007). Therefore, central bankers heavily rely on inflation forecasts to formulate monetary policies, in the pursuit of controlling inflation and achieving a stable price. To provide good forecasts of the inflation rate, economists have studied inflation from various perspectives. We will not attempt to provide even a partial review of a huge literature. Thus, in

this section, we focus on reviewing some main issues raised by economists regarding inflation modelling. By doing so, we aim to provide the readers with a general idea of the framework in inflation research. As a result, we hope the brief summarizations could stimulate readers' interest in this challenging research area. We then summarize some important references in the literature which interested readers could refer to.

The following topics are the current focus of research: The first focus is, by allowing for learning in macroeconomics, to model the dynamics of the inflation rate of inflation expectations with key economic variables, such as the unemployment rate and output. The second focus is to model inflation expectations, especially anchored inflation expectations. To do so, we have to model inflation uncertainties, as well as the impacts of the shocks on inflation. The third focus is modelling the underlying processes of the core and headline inflation rates, we could then evaluate how responsive inflation is to the food and energy shocks. Last but not least, price movements on both the aggregated level and disaggregated levels is of interest. By looking into the disaggregated price indices, we can investigate how inflation expectations change and how the pricing decisions are made. Using the disaggregated data, we may identify and forecast the weights of the various economic price indices according to our perceptions of the whole economy. Thereby, a summarization of the disaggregated data could provide an insight of the inflation behaviour on an aggregated level.

Further to the interesting topics above, some mysterious historical events have also drawn great attention since we believe learning from the past may help to reduce the likelihood of the same things to happen in the future. The following question is raised in the literature: Is inflation easier or harder to forecast nowadays (see Stock and Watson, 2007)?

Next, we provide some references in the literature with respect to the inflation rates modelling. In the literature, inflation rates models fall into two genres: (1) The first category includes the models based on the new Keynesian Phillips curve and the dynamic stochastic general equilibrium (DSGE) models. These models are normally applied to project the core inflation at longer-term horizons (Bernanke, 2007). (2) The second category consists of purely statistical models. Based on these statistical models, we project the near-term inflation using time series analysis techniques. Although this chapter focuses on univariate time series based on a statistical model, in the following, we provide some references to modelling of the inflation rate from various economic variables first, and some references on the statistical modelling regarding the inflation time series second.

The relationship between the inflation rate (also inflation expectations) and the

key economic variables, such as nominal interest rates, the unemployment rate, the openness of an economy, the exchange rate, and the tax system, have been profoundly studied. A recent article from Edward (2006) researches whether the exchange rates should be considered under the inflation targeting policy. The literature in this regards is too large to survey comprehensively.

Various statistical models have been applied to fit the inflation data sets. A very recent article from Watson and Stock (2007) applies a time varying trend-cycle (TC) model for the US inflation rates. Forecasting results from this time varying TC model provides smaller forecast errors comparing to the AR (4) model. They suggest that the univariate inflation process can be well described by a trend-cycle model with stochastic volatility or, equivalently, an integrated moving average process with time-varying parameters. Baillie et al.(1996) applied an ARFIMA-GARCH model to test the Friedman hypothesis. Kim (1993) applied a Markov-Switching model to investigate the inflation and inflation uncertainties. Cogley and Sargent (2001, 2005) applied a time varying AR model to investigate the inflation persistence. Many other statistical models, such as the State Space model, the simple AR (p) model, the Integrated Moving Average (IMA) model, and the cyclical models, have been used to model the inflation rate. From the perspective of purely statistical models, how to model the underlying process of the inflation and filter out the movements of the unobservable components, clearly, is an area that needs a lot of research.

4.1.2 Motivating the Stationary Bilinear Model

After a review of inflation modelling at a high level of abstraction, in this section, we introduce a purely statistical univariate nonlinear model, the Stationary Bilinear (SB) model, specified as the following:

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t \tag{4.1}$$

where b is the bilinear term and $\varepsilon_t \sim i.i.d.f_N\left(0, \sigma_\varepsilon^2\right)$.

First, we review the development of the Bilinear class of models and its extensions. Then, we look into the model specification and elaborate the feasibility of applying the SB model to the inflation series.

The bilinear models were first proposed by Granger and Anderson (1978). The class of bilinear process is found to be able to "approximate any nonlinear model to an arbitrary degree of accuracy over a finite time interval" (see Lane, Peel and Raeburn 1996). Brunner and Hess (1995) point out that this capacity of to approximate any

well-behaved nonlinear relationship is analogous to the ability of an ARMA model of approximating well-behaved linear relationships. Because of the capability of bilinear models to accommodate a wide variety of nonlinear behaviours, the bilinear class of models are interesting to many researchers. Further research of special cases in the bilinear class of models as carried out by Rao (1981), Quinn(1982), Hannan (1982), Martins (1997) and Igloi and Terdik (1999). The bilinear process is also studied under special circumstances, such as when extreme events exist. Davis and Resnick (1996) provided the asymptotic properties of a bilinear process and analyzed the stationary bilinear process generated from heavy-tailed noise variables. Other extensions of the bilinear process can be found in Giraitis and Surgailis (2002), where the properties of an ARCH-type bilinear models have been investigated. In Francq et al. (2008), the feasibility of the Dickey-Fuller and the PP test for a class of stochastic unit root bilinear process has been studied.

Since the bilinear models are simple non-linear extensions of the linear models and bilinear models can approximate most nonlinearities, bilinear models have been successfully applied to analyze macroeconomic and financial series in the economics context.

Byers and Peel (1995) use a bilinear-QARCH model to fit the inter-war exchange rate data. They found empirical evidence of bilinearity in both mean and variance. Peel and Davidson (1998) propose a Bilinear Error Correction Mechanism (BECM) for "data variables displaying abrupt changes". For an empirical illustration, they apply the BECM of an annual data on the real consumer's expenditure and the real gross national product in the UK. Their results show that the BECM is superior to the alternatives. Maravall (1983) used the bilinear model with an application of the supply of Spanish currency. He found that the application of the bilinear models for forecasting nonlinear processes demonstrates improvement over the ARIMA forecasts. Weiss (1986) stresses that bilinearity was important in explaining the behaviour of the sunspot data, the Canadian Lynx population data, and the stock price data. In addition, he emphasized that ignoring the possibility of bilinearity could lead to incorrect inference about conditional heteroscedasticity. Charemza et al. (2005) propose a Unit Root Bilinear process and a two-step testing procedure for detecting the bilinearity. They applied 65 countries stock market indices and they found strong evidence of the presence of Unit Root Bilinearity in a large number of mature and emerging markets. Charemza et al. (2006) use a bilinear model to evaluate the non-systematic part of the inflation, which is defined as the difference between the headline inflation and the core inflation. They applied 141 countries' monthly inflation rates and found the max-

imum admissible forecast horizons[1] in the developed countries are shorter than those in the developing countries. Other reviews of the bilinear model with an engineering perspective can be found in the work of Meddeb, Tourneret and Castanie (1998).

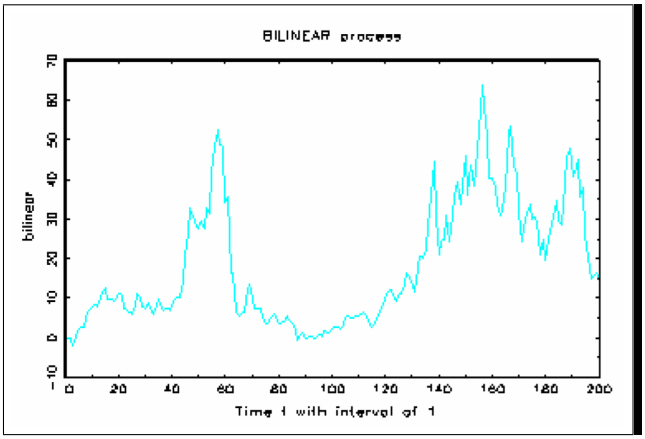


Figure 4-1A. Simulate Data series from a Unit Root Bilinear DGP

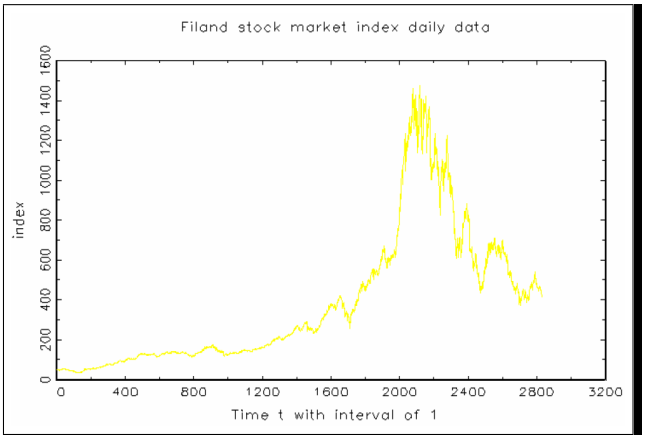


Figure 4-1B. Series Plot of Finland Stock Market Index: Daily Data

We simulated a data series from a Unit Root Bilinear DGP, and compare it with a daily Finland stock market index data. From a simple visual inspection, we found a lot of similarities between the simulated data and the real life data, see Fig (4-1A and 4-1B).

[1]The misspecifications in the forecasting will is created because of the existence of the bilinear component. The distorted forecast period depends on the forecast horizons. In Charemza et.al (2006), the maximum admissible forecast horizon is defined as the longest period, or the longest forecast horizon, that the distortions in the forcast are not substantial.

The bilinear model has good potential for being applied to macroeconomic and financial data, a better forecasting capacity than the ARIMA model (see Maravall, 1983) and can mimic the movements of the real life data to a certain extent. We are inspired to evaluate a first order Stationary Bilinear (SB) model, a special case of the bilinear class model, in a Bayesian context. The SB model is specified in Equation (4.1). Granger and Anderson (1978) point out, the necessary stationary restriction for a bilinear process is

$$a^2 + b^2\sigma_\varepsilon^2 < 1 \tag{4.2}$$

where a, $b \neq 0$. This necessary restriction together with three other necessary restrictions meet the assumption that the first four moments of $\{y_t\}_{t=1}^T$ are finite (see Sesay and Rao, 1988, Kim et al., 1990). The other three necessary restrictions are listed as follows:

$$|a| < 1 \tag{4.3}$$

$$\left|a^3 + 3ab^2\sigma_\varepsilon^2\right| < 1 \tag{4.4}$$

$$a^4 + 6a^2b^2\sigma_\varepsilon^2 + 3b^4\sigma_\varepsilon^4 < 1 \tag{4.5}$$

Unlike the Unit Root Bilinear (URB) process in Charemza et al. (2005)[2], in this chapter, a and b in the SB model jointly meet the stationary condition $a^2 + b^2\sigma_\varepsilon^2 < 1$. Since the AR coefficient $a + b\varepsilon_{t-1}$ is not independent on the error term ε_{t-1}, a nonzero b would induce a time varying parameter $a + b\varepsilon_{t-1}$ that changes correspondingly to the lagged shocks ε_{t-1}. Therefore, the series' persistence will not only depend on the long-range shocks, but will also depend on the magnitude and the sign of the shocks. Charemza et al. (2005) use stock prices to interpret the economic sense of b. They state that b "reflects the extent to which consumers' decisions regarding financial market transactions are affected by previous mistakes regarding share prices". According to Equation (4.1), one-step ahead forecast of y_{t+1} will be

$$E\left(y_{t+1}\right) = \left(a + b\varepsilon_t\right) y_t$$

which indicates that a big shock (extreme event) in the system may induce a jump in the series' persistence. Moreover, all the past shocks $\{\varepsilon_i\}_{i=1}^t$ will inevitably be propagated via Equation (4.1) to affect $E\left(y_{t+1}\right)$. If the near-term forecast $E\left(y_{t+1}\right)$ is

[2]In the URB model, a is assumed to equal to one. Then, a two-step testing procedure is proceeded to test the significance in the bilinear term b.

constructed in a way such that the bilinear term b and the contingency ε_t are neglected, the forecast will be severely distorted[3].

Although the bilinear process has great potential for empirical applications, the estimations of a bilinear process could be problematic. Brunner and Hess (1995) point out when the bilinear model is close to violating at least one of the four moment conditions in Equation (4.2, 4.3, 4.4, and 4.5), the expected likelihood function becomes bimodal. In such a situation, the true optimum is characterized by a long narrow spike and the conventional optimization routine would frequently miss this global optimum. As a consequence, it may not be appropriate to fit the bilinear model to the time series data following the standard model fitting procedure. Moreover, "the classical inference tests cannot be used for the bilinear models...and the t-statistic has very undesirable properties, especially when the model is close to violating its moment conditions" (Brunner and Hess, 1995). Brunner and Hess (1995) evaluate the Maximum likelihood estimates from the NPSOL[4] optimization algorithm. They found that when the moment conditions in Equation (4.2, 4.3, 4.4, and 4.5) are easily satisfied, the estimates of a and b converge to the true value as the sample size increases. However, in the cases that the model's moment conditions are almost violated, the maximum likelihood estimates have very poor properties, such as significant bias. In addition, "as the bilinear term becomes larger, the omitted variable bias introduced by ignoring the bilinear term causes the autoregressive parameter a to be substantially over-estimated" (Brunner and Hess, 1995). Based on above problems in the Maximum Likelihood Estimation (MLE) method for the bilinear model, we are inspired to evaluate the bilinear model in a Bayesian context.

Hristova (2005) applies the MLE method for estimating the coefficient in a Unit Root Bilinear process, in which the stationary condition in Equation (4.2) is broken. She maximizes the likelihood function by switching between a Newton-Raphson (NR) algorithm and a Simulated Annealing (SA) algorithm, in case that NR algorithm fails. Hristova (2005) claims that the SA algorithm could overcome the problems caused by bimodality in the expected likelihood function, and find the global optimum or a good,

[3]If $b = 0$, the SB model will be an AR(1) model:

$$y_t = ay_{t-1} + \varepsilon_t$$

and the one-step ahead forecast is:

$$E(y_{t+1}) = ay_t$$

[4]An algorithm invented by Philip Gill, Water Murray, Michael Saounders and Margaret Wright. This algorithm is to solve constrained optimization problems.

near optimal local optimum.

Motivated by the fact that the SB model has a great potential for being applied in economics or finance on the one hand, yet some potential estimation problems exist on the other hand, we focus on developing a posterior simulator for the SB model in the Bayesian context, in section 4.2 and section 4.3. To compare the estimation efficiency between the Bayesian MCMC and the classical MLE, we compare the Root Mean Square Errors (RMSE) of the estimates using 100 simulated series with various sample sizes, in section 4.4.1 and 4.4.2. In section 4.4.3, we evaluate the validity of the conventional ADF test and PP test under the existence of a bilinear process. For an empirical illustration, in section 4.5, we applied the UK's quarterly inflation rates series.

4.2 Bayesian inference of the SB Model

Following the Bayesian reasoning, we firstly derived the posterior conditionals from the likelihood function and the priors in section 4.2.1. Secondly, in section 4.2.2, we conducted two experiments to illustrate what forms of priors can be elicited.

4.2.1 Posterior Conditionals

We start by introducing some notations: a time series with a sample size of n is denoted as $y = (y_1, \cdots, y_n)'$, in which we assume the first observation y_1 as the initial value. The error disturbances ε_t for $t = 1, \cdots, n$ are $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$, where the initial disturbance value $\varepsilon_1 = 0$. The parameters of interest are $\theta = (a, b, h_\varepsilon)'$ in this SB model, and the error precision $h_\varepsilon = \sigma_\varepsilon^{-2}$. Taking Equation (4.1) recursively by

$$\varepsilon_{t-1} = y_{t-1} - (a + b\varepsilon_{t-2})\, y_{t-2}$$

we can express y_t as a function of F_{t-1}[5].

For $t = 3, \cdots, n$

$$y_t = a f_1(t, b) + f_2(t, b) + \varepsilon_t \tag{4.6}$$

For $t = 2$, $f_2(t, b) = 0$ and we have

$$y_t = a f_1(t, b) + \varepsilon_t \tag{4.7}$$

[5]Refer Appendix 4.A for detailed recursive derivations.

where

$$f_1(t,b) = \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right]$$

$$f_2(t,b) = \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right]$$

According to Equation (4.7) and Equation (4.6), the likelihood function $p(y_t \mid \theta, F_{t-1})$ is the following:

$$p(y_t \mid \theta, F_{t-1}) = \frac{h_\varepsilon^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \exp\left\{ -\frac{h_\varepsilon}{2} \left[y_t - a f_1(t,b) - f_2(t,b) \right]^2 \right\},$$

which leads to a joint likelihood function of $p(y \mid \theta)$ expressed as follows:

$$p(y \mid \theta) = \frac{h_\varepsilon^{\frac{N-1}{2}}}{(2\pi)^{\frac{N-1}{2}}} \exp\left\{ -\frac{h_\varepsilon}{2} \sum_{t=2}^{N} \left[y_t - a f_1(t,b) - f_2(t,b) \right]^2 \right\} \tag{4.8}$$

According to the necessary stationary restriction in Equation (4.2), the joint prior $p(a, b, h_\varepsilon)$ is:

$$p(a, b, h_\varepsilon) = p(b \mid h_\varepsilon, a)\, p(h_\varepsilon \mid a)\, p(a)\, 1\left(a^2 + b^2 \sigma_\varepsilon^2 < 1\right) \tag{4.9}$$

1. **The Unconditional Prior of** a

According to the necessary restriction in (4.3), it is reasonable to have a prior belief that a follows a truncated normal distribution within the unity region. The variance of a, $\underline{V}_a$ can be chosen at different values according to our "priori" belief, as to the way we expect a to vary. Then, we have the prior of a specified as $a \sim f_N\left(\underline{\mu}_a, \underline{V}_a\right) 1(|a| < 1)$, with a prior density as

$$p(a) = \frac{1}{\sqrt{2\pi \underline{V}_a}} \exp\left\{ -\frac{\left(a - \underline{\mu}_a\right)^2}{2\underline{V}_a} \right\} \cdot 1(|a| < 1) \tag{4.10}$$

where $\underline{\mu}_a$ is the prior mean and $\underline{V}_a$ is the prior variance.

For model comparison reasons, we would like to have a proper prior. Hence, a truncated proper prior of a can be obtained by dividing the improper prior's density with a normalizing constant $\Pr(1 \,||a| < 1|)$. Therefore, the proper prior of a is as follows:

$$p(a) \propto \frac{1}{\Pr\left(1\left||a|<1\right|\right)\sqrt{2\pi \underline{V}_a}} \exp\left\{-\frac{\left(a-\underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} \cdot 1\left(|a|<1\right)$$

where the normalizing constant $\Pr\left(1\left||a|<1\right|\right)$ is:

$$\Pr\left(1\left||a|<1\right|\right) = \Phi_N\left(\frac{1-\underline{\mu}_a}{\sqrt{\underline{V}_a}}\right) - \Phi_N\left(\frac{-1-\underline{\mu}_a}{\sqrt{\underline{V}_a}}\right)$$

2. **A Conditional Prior of** h_ε

The conditional prior for h_ε follows a Gamma distribution $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) 1(a^2 + b^2/h_\varepsilon < 1)$ and b is uniformly distributed within a region such that the stationary condition in Equation (4.2) can be satisfied. The joint prior of h_ε and b conditional on a can be expressed as follows:

$$p\left(h_\varepsilon, b \mid a\right) \propto \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) 1(a^2 + b^2/h_\varepsilon < 1) \tag{4.11}$$

The conditional prior for $p\left(h_\varepsilon \mid a\right)$ then can be obtained by marginalizing over b in Equation (4.11) in the following:

$$p\left(h_\varepsilon \mid a\right) \propto \int p\left(h_\varepsilon, b \mid a\right) db$$

According to the stability condition, b is within a range specified as below:

$$-\sqrt{\left(1-a^2\right)h_\varepsilon} < b < \sqrt{\left(1-a^2\right)h_\varepsilon}$$

Since Equation (4.11) is not a function of b, the conditional prior of h_ε, $p\left(h_\varepsilon \mid a\right)$ can be obtained as the following:

$$\begin{aligned} p\left(h_\varepsilon \mid a\right) &\propto \int p\left(h_\varepsilon, b \mid a\right) 1(a^2 + b^2/h_\varepsilon < 1) db \\ &\propto p\left(h_\varepsilon, b \mid a\right) \int_{-\sqrt{(1-a^2)h_\varepsilon}}^{\sqrt{(1-a^2)h_\varepsilon}} db \\ &\propto \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) \cdot 2\sqrt{\left(1-a^2\right)h_\varepsilon} \end{aligned} \tag{4.12}$$

3. A **Conditional Prior of** b

We presume that b is uniformly distributed within a region, where the moment conditions are met. Then, a proper conditional prior $p(b \mid h_\varepsilon, a)$ for b is the following:

$$p(b \mid h_\varepsilon, a) \propto \frac{1}{2\sqrt{(1-a^2)\, h_\varepsilon}} \cdot 1\left[|b| < \sqrt{(1-a^2)\, h_\varepsilon}\right] \tag{4.13}$$

With the derived likelihood in (4.8) and the joint prior in (4.9), we are able to derive a joint posterior according to the Bayes Theorem,

$$p(a, b, h_\varepsilon \mid y) \propto p(y \mid a, b, h_\varepsilon)\, p(a, b, h_\varepsilon)$$

Although we have the joint posterior for a, b and h_ε, the marginal posterior conditionals of each individual parameter are ambiguous. Next, we demonstrate the derived marginal posterior conditionals of a, b and h_ε and their detailed derivations can be found in Appendix 4.B. With the marginal posterior densities, the MCMC algorithms can then be implemented to simulate the marginal posterior distributions. The marginal posterior conditionals are as follows:

1. **The posterior conditional of** a

The posterior conditional of a is $a \sim f_N\left(\overline{\mu}_a, \overline{V}_a\right) 1\left(|a| < \sqrt{1 - b^2/h_\varepsilon}\right)$, a truncated normal distribution with a variance $\overline{V}_a$

$$\overline{V}_a = \left[\frac{\sum_{t=2}^{N} f_1^2(t, b)}{\sigma_\varepsilon^2} + \frac{1}{\underline{V}_a}\right]^{-1}$$

and a mean $\overline{\mu}_a$

$$\overline{\mu}_a = \overline{V}_a \cdot \left[h_\varepsilon \sum_{t=2}^{N} f_1(t, b)\left[y_t - f_2(t, b)\right] + \frac{\underline{\mu}_a}{\underline{V}_a}\right] \tag{4.14}$$

2. **The posterior of the error precision** h_ε

The posterior conditionals for h_ε follows a truncated Gamma distribution with a density $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right) 1\left[h_\varepsilon > \frac{b^2}{1-a^2}\right]$, where

$$\overline{\alpha}_\varepsilon = \underline{\alpha}_\varepsilon + \frac{N-1}{2}$$

and

$$\overline{\beta}_{\varepsilon} = \left[\frac{1}{\underline{\beta}_{\varepsilon}} + \frac{1}{2} \sum_{t=2}^{N} \left[y_t - a f_1 \left(t, b \right) - f_2(t, b) \right]^2 \right]^{-1}$$

If $\frac{b^2}{1-a^2}$ is large, extreme values of h_{ε} have to be sampled from the upper tail of the Gamma distribution. For simplicity, we approximate the tail of $f_{\Gamma\varepsilon}$ with an exponential distribution

$$f_{\exp}\left(x \right) = \lambda \exp\left(-x\lambda \right),$$

where $\lambda = \frac{b^2}{1-a^2}$. A simple Inverse c.d.f Sampling with an exponential rejection sampling method is applied for generating random draws of h_{ε}. This will be discussed in detail in the section of methodology issues, in section 4.3.1.

3. **The full conditional density of** b

The value of b depends on the values of a and h_{ε}. The posterior conditional of b is:

$$p\left(b|y, a, h_{\varepsilon} \right) \propto 1 \left[|b| < \sqrt{\left(1 - a^2 \right) h_{\varepsilon}} \right] \cdot \exp \left\{ -\frac{h_{\varepsilon}}{2} \sum_{t=2}^{N} \left[y_t - a f_1 \left(t, b \right) - f_2(t, b) \right]^2 \right\} \tag{4.15}$$

Since b is bounded within a region to meet the stationarity requirement, and b does not follow a standard distribution form, from which we can generate random samples, we apply the Griddy-Gibbs sampling method to get random draws of b with this non-standard density. Refer to chapter 2.3.2 for the reviews of Griddy-Gibbs sampling method. Based on the available posterior conditionals of a, b and h_{ε}, a straight forward Gibbs samplings incorporated with a Griddy-Gibbs sampling algorithm can be carried out to simulate the posteriors of a, b and h_{ε}.

4.2.2 Prior Elicitations

To specify an appropriate prior that adequately reflects the available information is important. In this chapter, since investigating the UK's inflation rates will be our main focus, we have to elicit appropriate prior parameters that are able to mimic the patterns of the real life data. This prior elicitation procedure is the so-called "data based prior elicitations".

Imagine a series of T observations, we can count the proportion of the observations jumping above x. Considering the UK's inflation rates, the probability of this propor-

tion then should be realistic to a certain extent. Thus, if this proportion is denoted as *prop*, we want to ask what is the probability that *prop* exceeds s?

To answer the above question, with the chosen prior distributions, we randomly generate $2,000$ vectors of numbers, $\left\{\theta_n^* = (a_n^*, b_n^*, h_n^*)'\right\}_{n=1}^{2000}$. Then, with these vectors, we generate $2,000$ series' with $T = 200$ observations following the SB DGP. Within each single series, the proportion (denoted as *prop*) of y_t goes beyond 40 and below -40 can be counted, denoted as $prop\left(|y_t| > x\right)$, where $x = 40$. If this proportion is greater than 0.5 in this particular simulated series, denoted as $prop\left(|y_t| > x\right) > s$, where $s = 0.5$, we keep a record of this series. The notation $prop\left(|y_t| > 40\right) > 0.5$ actually stands for a category, which indicates that within one particular series, more than 50% of the observations goes beyond 40 or below -40. Then, we examine through all the simulated data series' by counting the number of series, denoted as n, out of $2,000$ that fall into this category. Therefore, the percentage $n/2000$ can be an approximation of the probability that a DGP consistently (more than half of the time) generates observations beyond 40 or below -40, denoted as $\Pr\left[prop\left(|y_t| > 40\right) > 0.5\right]$.

If $\Pr\left[prop\left(|y_t| > 40\right) > 0.5\right]$ is high, it means that a great number of the series out of the total $2,000$ are having large proportions (50%) of observations beyond a certain level ($x = 40$). Consequently, it indicates that the prior is not a reasonable prior because with the values θ_n^* generated from the prior, the SB DGP is incapable to mimic the pattern of the real life data. Thus it makes little sense to use this sort of elicited prior to evaluate the UK inflation rate.

Similarly, we apply the exact same procedure to elicit the prior in the RW model. Next, we illustrate how to choose the parameter values in the prior via simulations in more details.

1. **Choose Values in the Prior of Error Precision h_ε**

In the Gamma prior of $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) 1(a^2 + b^2/h_\varepsilon < 1)$, $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ need to be selected with caution. We want to select the appropriate priors that generate reasonable values of h_ε. After an inspection of the monthly UK inflation rates, within 585 observations, 92 observations are above 10 ($prop\left(|y_t| > 10\right) > 0.15$) and 17 observations are above 20 ($prop\left(|y_t| > 20\right) > 0.02$). Thus, we should select a prior, with which the simulated series could illustrate such properties.

In the prior $a \sim f_N\left(\underline{\mu}_a, \underline{V}_a\right) 1\left(|a| < 1\right)$, $\underline{\mu}_a$ is selected as 1 and $\underline{V}_a$ is selected as 100^2. It indicates a belief of a being uniformly distributed within a truncated region. Then, we select 9 different combinations of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ in the prior $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) 1(a^2 + b^2/h_\varepsilon < 1)$ and simulate $2,000$ series from each of the priors. The

characteristics of simulated data from SB DGP are shown in Table (4.1).

With the same method, we use the simulate series to elicit the prior for $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ in the RW model. Table (4.2) illustrate the characteristics of the simulated data series with different values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$.

From Table (4.1), We select $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$ in $f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$as a reasonable prior for h_ε with the SB DGP[6]. With this prior, we get the highest probability (0.3335) that more than 10% of the observations are larger than 10, which can be expressed as $\Pr\left[prop\left(|y_t| > 10\right) > 0.1\right] = 0.3335$ with $x = 10$ and $s = 0.1$. With this prior, the probability $\Pr\left[prop\left(|y_t| > 20\right) > 0.02\right]$, which indicates that around 2% observations are larger than 20, is also the highest compared with other prior choices. The SB DGP is more likely to mimic the movements in the UK inflation rates series under this elicited prior. Therefore, other choices of priors will be less appropriate. For instance, if we choose $\underline{\alpha}_\varepsilon = 10$ and $\underline{\beta}_\varepsilon = 0.1$, from Table (4.1) we can read $\Pr\left[prop\left(|y_t| > 10\right) > 0.1\right] = 0.0400$. This indicates that the probability of more than 10% of the samples being larger than 10 is only 0.0400. In other words, in most of the simulated series', large number of observations are less than 10, which makes it seeming unlikely that the SB DGP represents the UK inflation according to the UK's historical inflation data.

For the same reasons, we select the same values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ in the RW as those in the SB for the prior $h_\varepsilon \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$.

2. **Choose Values in the Prior of a in the SB**

Following the same prior eliciting procedure as that of σ_ε^2, we have another look at the prior of $a \sim f_N\left(\underline{\mu}_a, \underline{V}_a\right) 1\left(|a| < 1\right)$ to see if setting $\underline{V}_a = 100^2$ is appropriate. The value of $\underline{\mu}_a$ is chosen as 1. There are two types of value for $\underline{V}_a$ that can be chosen: one is a "loose" prior with very large value of $\underline{V}_a$, and the other is a "tight" prior with a very small value of $\underline{V}_a$. If we choose $\underline{V}_a$ as a large value, it indicates a belief of a being approximately uniformly distributed within the truncated region. If we choose $\underline{V}_a$ as a small value, it indicates we believe a to vary within a very tight region around the mean $\underline{\mu}_a$ where the stationary condition is satisfied.

In the SB DGP, with the elicited prior of $h_\varepsilon \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) 1(a^2 + b^2/h_\varepsilon < 1)$, where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$, we select a range of values of $\underline{V}_a$ from 0.01 to 100^2 to illustrate the characteristics of the simulated data. Based on Table (4.3), we are able to choose a value of $\underline{V}_a$ to elicit an appropriate prior of a.

[6]The inverse of h_ε, $1/h_\varepsilon$ follows an Inverse-Gamma $f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$.

Table 4.1: Using Simulation to Elicit Priors of SB

$prob[\lvert prop(y > x)\rvert > s]$		$SBDGP$				
$\alpha = 10$						
		s/x	10	20	30	40	50
	$\beta = 0.1$	0.1	0.0400	0.0065	0.0035	0.0030	0
		0.3	0.0065	0.0020	0.0010	0	0
		0.5	0.0025	0.0010	0	0	0
	$\beta = 0.2$	0.1	0.0135	0.0025	0.0005	0.0005	0.0005
		0.3	0.0040	0.0005	0.0005	0	0
		0.5	0.0025	0.0005	0	0	0
	$\beta = 1$	0.1	0.0020	0.0010	0	0	0
		0.3	0	0	0	0	0
		0.5	0	0	0	0	0
$\alpha = 5$							
	$\beta = 0.1$	0.1	0.1085	0.0185	0.0075	0.0050	0.0035
		0.3	0.0165	0.0050	0.0030	0.0015	0.0015
		0.5	0.0075	0.0025	0.0015	0.0010	0
	$\beta = 0.2$	0.1	0.0365	0.0075	0.0025	0.0005	0
		0.3	0.0080	0.0010	0	0	0
		0.5	0.0020	0	0	0	0
	$\beta = 1$	0.1	0.0015	0.0005	0	0	0
		0.3	0.0005	0	0	0	0
		0.5	0	0	0	0	0
$\alpha = 2$							
	$\beta = 0.1$	0.1	0.3335	0.0905	0.0360	0.0165	0.0100
		0.3	0.0900	0.0160	0.0090	0.0060	0.0030
		0.5	0.0275	0.0070	0.0035	0.0020	0.0015
	$\beta = 0.2$	0.1	0.1645	0.0310	0.0125	0.0070	0.0035
		0.3	0.0330	0.0045	0.0015	0	0
		0.5	0.0065	0.0005	0	0	0
	$\beta = 1$	0.1	0.0230	0.0020	0.0010	0	0
		0.3	0.0015	0.0005	0	0	0
		0.5	0.0005	0	0	0	0

Table 4.2: Using Simulation to Elicit Priors of RW

$prob[\lvert prop(y > x)\rvert > s]$		$RWDGP$				
$\alpha = 10$						
		s/x	10	20	30	40	50
	$\beta = 0.1$	0.1	0.6285	0.1785	0.0445	0.0105	0.0020
		0.3	0.4090	0.0885	0.0150	0.0020	0
		0.5	0.2375	0.0335	0.0020	0.0005	0
	$\beta = 0.2$	0.1	0.3995	0.0460	0.0050	0.0020	0.0005
		0.3	0.2130	0.0170	0.0025	0	0
		0.5	0.1065	0.0055	0.0005	0	0
	$\beta = 1$	0.1	0.0325	0.0010	0	0	0
		0.3	0.0105	0	0	0	0
		0.5	0.0035	0	0	0	0
$\alpha = 5$							
	$\beta = 0.1$	0.1	0.8570	0.4165	0.1750	0.0680	0.0290
		0.3	0.6310	0.2325	0.0845	0.0265	0.0135
		0.5	0.4055	0.1230	0.0330	0.0120	0.0030
	$\beta = 0.2$	0.1	0.6585	0.2120	0.0520	0.0105	0.0030
		0.3	0.4335	0.1070	0.0175	0.0040	0.0010
		0.5	0.2665	0.0370	0.0040	0.0005	0
	$\beta = 1$	0.1	0.1550	0.0075	0.0005	0	0
		0.3	0.0725	0.0030	0	0	0
			0.0280	0.0010	0	0	0
$\alpha = 2$							
	$\beta = 0.1$	0.1	0.9755	0.7635	0.5310	0.3545	0.2370
		0.3	0.8745	0.5555	0.3360	0.2010	0.1325
		0.5	0.6870	0.3635	0.1935	0.1130	0.0640
	$\beta = 0.2$	0.1	0.9005	0.5590	0.3045	0.1735	0.0940
		0.3	0.7275	0.3605	0.1740	0.0830	0.0450
		0.5	0.5205	0.2065	0.0840	0.0415	0.0260
	$\beta = 1$	0.1	0.4680	0.1245	0.0355	0.0145	0.0050
		0.3	0.2990	0.0645	0.0175	0.0065	0.0020
		0.5	0.1735	0.0295	0.0075	0.0015	0.0010

Table 4.3: Simulation to Elicit Prior of Parameter a in SB

Va	s/x	5	10	20	30	40
0.01	0.1	0.5975	$0.205^{\dagger}$	0.038	0.0145	0.007
	0.3	0.194	0.0385	0.0065	0.002	0.0005
	0.5	0.0595	0.0105	0	0	0
0.05	0.1	0.467	0.1285	0.03	0.009	0.0045
	0.3	0.1105	0.022	0.0045	0.002	0
	0.5	0.034	0.0075	0.002	0	0
01	0.1	0.3525	0.0895	0.016	0.0075	0.003
	0.3	0.072	0.0125	0.0025	0.001	0.001
	0.5	0.019	0.004	0.0015	0.001	0.0005
0.5	0.1	0.2145	0.047	0.0115	0.003	0.001
	0.3	0.038	0.0105	0.001	0	0
	0.5	0.0125	0.005	0.0005	0	0
1	0.1	0.1775	0.038	0.008	0.0035	0.001
	0.3	0.03	0.008	0.0025	0.001	0.0005
	0.5	0.0105	0.0025	0.001	0.0005	0
10	0.1	0.1425	0.0255	0.003	0.001	0.0005
	0.3	0.023	0.006	0.0005	0	0
	0.5	0.009	0.002	0	0	0
100	0.1	0.131	0.025	0.0035	0.0015	0.001
	0.3	0.026	0.0065	0.002	0.0005	0
	0.5	0.0105	0.0035	0.0005	0	0
100^2	0.1	0.15	$0.0315^{\ddagger}$	0.0085	0.0035	0.002
	0.3	0.0285	0.0085	0.0025	0.001	0
	0.5	0.011	0.003	0.002	0	0

Table 4.4: Summarized Table of Elicited Priors in SB and RW

SB	RW
$h_\varepsilon \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) 1(a^2 + b^2/h_\varepsilon < 1)$	$h_\varepsilon \sim f_{\Gamma_\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$
$\underline{\alpha}_\varepsilon = 2$, $\underline{\beta}_\varepsilon = 0.1$	$\underline{\alpha}_\varepsilon = 2$, $\underline{\beta}_\varepsilon = 0.1$
$a \sim f_N\left(\underline{\mu}_a, \underline{V}_a\right) 1\left(\lvert a\rvert < 1\right)$	
$\underline{\mu}_a = 1$, $\underline{V}_a = \left\{\begin{smallmatrix} 0.01 \\ 100^2 \end{smallmatrix}\right\}$	

With a tight prior by choosing $\underline{V}_a = 0.01$, we have $\Pr\left[prop\left(|y_t| > 10\right) > 0.1\right] = 0.205$, which is the highest comparing with other choices. With a loose prior by choosing $\underline{V}_a = 100^2$, the probability is decreased to 0.0315. This indicates that if we choose $\underline{V}_a = 100^2$ in the prior of a, the probability of a series generating more than 10% of observations exceeding 10 is as low as 0.0315. With $\underline{V}_a = 100^2$, numbers generated from the prior distribution of a tend to be very small, which could easily satisfy the stationary conditions[7]. Therefore, most of the simulated series tend to be stationary with observations smaller than 5. As a consequence, the prior with $\underline{V}_a = 100^2$ may be less appropriate.

Based on the experiments, our first choice for a prior is $\underline{V}_a = 0.01$ or small values as a tight prior. This tight prior indicates a belief of high densities around the mean $\underline{\mu}_a = 1$, around which the stationary restriction is very volatile. However, as we would like to see how efficiently the Bayesian approach to estimating performs when the prior is elicited in such a way that a is believed to be a value not violating the stationary restriction with high probabilities, we also elicit $\underline{V}_a = 100^2$ and compare the estimation efficiency under these two distinct priors. Later on, in section 4.4.2, we evaluate the estimation efficiency under different elicited priors, with different sample sizes under the controlled settings with uses of the simulated artificial data sets. The selected priors are summarized in Table (4.4).

4.3 Methodology Issues

This section focuses on some substantial methodology issues regarding the estimation and the marginal likelihood evaluations of the SB model. Although we have provided a general review of the sampling and the marginal likelihood calculation algorithms in chapter 2.3 and 2.4, it is important to illustrate how to apply those algorithms in this

[7]This is because with $\underline{V}_a = 100^2$, a tend to be uniformly distributed within the truncated region where the moment conditions are satisfied.

particular SB model framework. In this regard, section 4.3.1 focuses on the sampling algorithms and section 4.3.2 introduces the marginal likelihood calculation in the SB model context.

4.3.1 Sampling Method for Estimation

This section introduces the sampling algorithms regarding each parameter of interest in the SB model. Four algorithms are used together to estimate the SB model. To sample a, we apply two methods: one is Algorithm 1(a), the Inversed Cumulative Density Function (c.d.f.) sampling method (Devroye, 1986), and the other is Algorithm 1(b), the Mixed Rejection sampling method (Geweke, 1991). Algorithm 2 applies the Griddy Gibbs sampling method to sample b, in which we set the grid number as 100. In Algorithm 3, the Mixed Rejection sampling method is applied again to get samples of h_ε. Algorithm 4 is a standard Gibbs sampler, which provides sequential draws of a, b and h_ε.

The following notations are employed in this section: The uniform distribution on the interval of a lower boundary and a upper boundary is denoted as $f_U[lo, up]$. The univariate truncated normal distribution is denoted as $f_N(\cdot)\,1(lo < x < up)$, with a density:

$$\frac{1}{[\Phi_N(up) - \Phi_N(lo)]} f_N(\cdot)$$

on the interval of $[lo, up]$ and 0 elsewhere. The truncated normal distribution is denoted as $f_\Gamma(\cdot)\,1(lo < x < up)$. If the truncated region lies in the distribution tails such that $lo = +\infty$ or $up = -\infty$, the density should be handled with special treatments. We will discuss in details in Algorithm 3.

Next, we review the sampling method involved in drawing random samples of a. The traditional rejection sampling method, where the random samples are dropped if the sampled draws are out of the truncated region, is very impractical because the probability of the acceptance is very unpredictable. For efficient sampling, we have to employ the inverse c.d.f. sampling method and the mixed rejection sampling method.

According to the posterior conditionals for a, a should be sampled from a truncated normal distribution $f_N(\overline{\mu}_a, \overline{V}_a)\,1\left(|a| < \sqrt{1 - b^2/h_\varepsilon}\right)$. The Inversed c.d.f. sampling procedure for a can be conducted as follows:

Algorithm 1(a): Inversed c.d.f. Sampling for a

1. Calculate the lower boundary and the upper boundary of the truncated region:

$\left(-\sqrt{1-b^2/h_\varepsilon},\ \sqrt{1-b^2/h_\varepsilon}\right)$. Hence, the corresponding boundaries in a $f_N(0,1)$ will be $lo = \frac{-\overline{\mu}_a - \sqrt{1-b^2/h_\varepsilon}}{\sqrt{V_a}}$ and $up = \frac{-\overline{\mu}_a + \sqrt{1-b^2/h_\varepsilon}}{\sqrt{V_a}}$.

2. Calculate the c.d.f. $\Phi_N(lo)$ and $\Phi_N(up)$

3. if we want to get a random draw a from a truncated normal distribution, $a \sim f_{TN}(lo, up)$, then

$$a = \Phi_N^{-1}(u)$$

$$u \sim f_U[\Phi_N(lo), \Phi_N(up)]$$

As pointed out by Geweke (1991), "this method requires the evaluation of one integral for each draw, and if the values of lo and up change with the draws, then three evaluations are required".

This sampling method requires the evaluation of the $\Phi_N(x)$, $\Phi_N(lo)$ and $\Phi_N(up)$. If any of the above values approximates to 0, or 1, the computation will take a long time. Geweke (1991) points out, if up is larger than 8 or lo is smaller than -8, the double precision implementation in most statistical softwares are unable to compute. I therefore apply the mixed rejection sampling method, which is proposed in Geweke (1991). This method is very efficient when we require draws from a truncated univariate normal distribution. Samples can be drawn as long as $|lo| \leq 35$, $|up| \leq 35$, when programmed in double precision $(64 - bit)$ floating point arithmetic.

The mixed rejection algorithm incorporates four different kinds of rejection sampling, depending on the value of lo and up. (1) Normal rejection sampling, x is drawn from $f_N(0,1)$ and accepted if $x \in [lo, up]$. (2) Half-normal rejection sampling, x is drawn from $f_N(0,1)$ and accepted if $|x| \in [lo, up]$, where ($lo \geq 0$). (3) Uniform rejection sampling, x is drawn from $f_U[lo, up]$, u is drawn independently from $f_U[0,1]$, and x is accepted if $u \leq f_N(x)/f_N(x^*)$, $\mathrm{x}^* = argmax\left[\Phi_N(x)\right]$[8]. (4) The Key point in the mixed rejection sampling is the use of an exponential rejection sampling. The motivation of applying the exponential rejection sampling is that if the truncated region is in the tails, where c.d.f. $\Phi_N(lo) \to 1$ and c.d.f. $\Phi_N(up) \to 1$(or $\Phi_N(lo) \to 0$ and $\Phi_N(up) \to 0$), the truncated normal distribution comes to resemble an exponential distribution, see Geweke (1991) for details. Hence, under any of the above conditions,

[8]In mathematics, $argmax$ stands for the argument of the maximum, that is to say, the value of the given argument for which the value of the given expression attains its maximum value. In this case, at x^*, $f_N(x)$ attains the maximum values within $[lo, up]$ region.

x should be drawn from $f_{\exp}(x)$, with a p.d.f. as:

$$f_{\exp}(x) = \lambda \exp(-x\lambda)$$

To obtain the highest acceptance probability, $\lambda = lo$ if $lo > 0$ and x is accepted if $x > lo$. $\lambda = up$ if $up < 0$ and x is accepted if $x < up$.

According to Geweke (1991), we have to employ four constants with this algorithm: $t_1 = 0.150$, $t_2 = 2.18$, $t_3 = 0.725$, $t_4 = 0.45$. These four values have been set through experimentations with computation time in Geweke (1991). The selected value is indicated when the constant is introduced. The sampling procedure depends on the relative configuration of lo and up. The mixed rejection sampling method for a is conducted as follows:

Algorithm 1(b): Mixed Rejection Sampling for a

1. The same as step 1 in **Algorithm 1(a)**.
2. Get a random draw of x
   (a) On (lo, ∞): normal rejection sampling if $lo \leq t_4$; exponential rejection sampling if $lo > t_4$
   (b) On (lo, up) if $0 \in [lo, up]$
      i. If $f_N(lo) \leq t_1$ or $f_N(up) \leq t_1$: normal rejection sampling
      ii. If $f_N(lo) > t_1$ or $f_N(up) \geq t_1$: uniform rejection sampling
   (c) On (lo, up) if $lo > 0$
      i. If $f_N(lo)/f_N(up) \leq t_2$: uniform rejection sampling
      ii. If $f_N(lo)/f_N(up) > t_1$ and $lo < t_3$: half normal rejection sampling
      iii. If $f_N(lo)/f_N(up) > t_1$ and $lo \geq t_3$: exponential rejection sampling.
   (d) The omitted cases, $(-\infty, up)$ and (lo, up) with $up < 0$, are symmetric to the case a and case c, respectively, and are treated in the same way.
3. Then, the random sample of a will be $a = \overline{\mu}_a + \sqrt{\overline{V}_a} \cdot x$

These two methods are tested through the experiments. The execution time of the Mixed Rejection method is smaller than the Inversed c.d.f. sampling method. For $10,000$ draws, the Mixed Rejection method took 920 seconds while the Inversed c.d.f. took $2,320$ seconds. Results show that the Mixed Rejection algorithm is 1.5 times faster than the Inversed c.d.f. algorithm.

Algorithm 2: Griddy Gibbs Sampling for b

1. Calculate the boundaries (lo, up) with $lo = -\sqrt{(1-a^2)\,h_\varepsilon}$ and $up = \sqrt{(1-a^2)\,h_\varepsilon}$ Then we divide the boundary with 100 grids and each step is defined as $step = (up - lo)/100$. Each grid point can be calculated as $b_i = lo + 0.5step \cdot i$, where $i = 1, \cdots, 100$.

2. Calculate $f(b_i) : i = 1, \ldots, 100$ according to the posterior density in Equation (4.15), $g(b_j) = \sum_{n=1}^{j} f(b_n)$ and

$$C(b_j) = \frac{g(b_j)}{\sum_{i=1}^{100} f(b_i)} : j = 1, \ldots, 100$$

1. Get a random draw from the uniform distribution $u \sim f_U(0, 1)$

2. Find the grid point b_j, where

$$b_j = C^{-1}(u)$$

Note that maybe only a small number of grid points within the boundaries contribute to the sum of the probabilities. In other words, the posterior density $f(b_i)$ may be sensitive to the value of b_i. Keeping in mind that the Griddy-Gibbs sampling method is actually approximating a continuous density with a number of discrete grids. If the grids are too coarse, we would inevitably lose the distribution on features of b. Therefore, the random draws of b using a coarse grid maybe less efficient. In this regard, we choose the grid number as 100.

Algorithm 3: Mixed Rejection Sampling for h_ε

Since we may have to draw extremely large values in the Gamma tails, a mixed rejection sampling method is also involved for h_ε.

1. According to the stationary condition, $h_\varepsilon > b^2/(1-a^2)$, where $lo = b^2/(1-a^2)$, we have to obtain the random draws from the upper tail of the Gamma distribution.

2. Calculate $\Phi_\Gamma(lo)$ with the corresponding posterior density $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right)$ and $\Phi_\Gamma = 1 - \Phi_\Gamma(lo)$.

   (a) If $\Phi_\Gamma \geq 0.05$, get a random draw from the uniform distribution $u \sim f_U(0, 1)$. A random draw of h_ε can be obtained using the Inversed c.d.f method

$$h_\varepsilon = \Phi_\Gamma^{-1}\left[\Phi_\Gamma(lo) + \Phi_\Gamma \cdot u\right]$$

(b) If $\Phi_\Gamma < 0.05$, we employ the exponential rejection method and the density in the Gamma tail is resembled with an exponential density

$$f_{\exp}(h_\varepsilon) = \lambda \exp(-h_\varepsilon \lambda)$$

where $\lambda = b^2/(1-a^2)$. Accept the draw h_ε if $h_\varepsilon > \lambda$.

Algorithm 4: Gibbs Sampling for a, b, h_ε

1. Give initial values to a, b, h_ε and the final results will not depend on the initial values.
2. Sample h_ε from $h_\varepsilon \mid a, b \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right) \cdot 1(A)$.
3. Sample a from $a \mid b, h_\varepsilon \sim f_N\left(\overline{\mu}_a, \overline{V}_a\right) \cdot 1(A)$ using Mixed rejection sampling
4. Sample b from $b \mid a, h_\varepsilon \sim f(b) \cdot 1(A)$ using the Griddy-Gibbs sampling, see chapter 2.2.3 for the reviews.

After taken S replications, with the first S_0 draws discarded, we can have the simulated distributions for a, b, h_ε. Then the estimates of the parameters will be the means of the posterior distributions and the convergence diagnostics can be conducted to examine the MCMC efficiency.

4.3.2 Model Comparison: Gelfand-Dey Method

In the SB model, since we know the forms of the exact likelihood function and the full p.d.f. of the proper priors, the Gelfand-Dey Method can be applied to evaluate the model's marginal likelihood. Refer to chapter 2.2.4 for reviews of the Gelfand-Dey Method.

Using the Gelfand-Dey method, the joint prior $p(a, b, h_\varepsilon)$ is evaluated at posterior draws $a^{(g)*}, b^{(g)*}$ and $h_\varepsilon^{(g)*}$,where $g = 1, \cdots, G$. If we denote $\theta^* = (a^*, b^*, h_\varepsilon^*)'$ and $\theta^{(g)*} = (a^{(g)*}, b^{(g)*}, h_\varepsilon^{(g)*})'$, the joint prior evaluated at each posterior draw is as the following

$$p\left(\theta^{(g)*}\right) = p\left(b^{(g)*} \mid h_\varepsilon^{(g)*}, a^{(g)*}\right) p\left(h_\varepsilon^{(g)*} \mid a^{(g)*}\right) p\left(a^{(g)*}\right) / C_t$$

where C_t is the integrating constant over the truncated region.

Algorithm 5: Marginal Likelihood Using the Gelfand-Dey Method

1. Get $\widehat{\theta}$ and $\widehat{\sum}$, where

$$\widehat{\theta} = \frac{1}{G}\sum_{g=1}^{G}\theta^{(g)*}$$

and

$$\widehat{\sum} = \frac{\theta'\theta}{G} - \widehat{\theta}\widehat{\theta}'$$

from the posterior simulator. Then, the truncated region is

$$\widehat{\Theta} = \left\{\theta : \left(\theta - \widehat{\theta}\right)' \widehat{\sum}^{-1} \left(\theta - \widehat{\theta}\right) \leq \chi^2_{1-p}(k)\right\}$$

p can selected as different values for the truncated tails. In our case, we selected a range of values that $p = [0.01, 0.02, 0.03]$. $k = 3$, which is the dimension of the parameter space in the SB model.

2. $f_{MN}(\theta)$ is the multivariate Normal density truncated to the region $\widehat{\Theta}$, for $\theta^{(g)*}$: $g = 1, \cdots, G$

$$f_{MN}(\theta^{g*}) = \frac{\frac{1}{p}}{(2\pi)^{\frac{k}{2}}} \left|\widehat{\sum}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta^{(g)*} - \widehat{\theta}\right)' \widehat{\sum}^{-1} \left(\theta^{(g)*} - \widehat{\theta}\right)\right] 1\left(\theta^{g*} \in \widehat{\Theta}\right)$$

If the condition $\left(\theta^{(g)*} - \widehat{\theta}\right)' \widehat{\sum}^{-1} \left(\theta^{(g)*} - \widehat{\theta}\right) \leq \chi^2_{1-p}(k)$ is not met, $f_{MN}\left(\theta^{(g)*}\right) = 0$.

3. Evaluate the likelihood ordinates at posterior draws $p\left(y|\theta^{(g)*}\right)$ according to Equation (4.8) and the prior ordinates at $p\left(\theta^{(g)*}\right)$ at $\theta^{(g)*}$ according to Equation (4.9).

4. Then, the marginal likelihood can be approximated as the following:

$$p(y \mid M_j) \propto E\left[\frac{f_{MN}(\theta^*)}{p(\theta^* \mid M_j)\, p(y \mid \theta^*, M_j)} \mid y, M_j\right]^{-1}$$

, where the standard MCMC convergence diagnostic tools can be applied.

4.4 Bayesian Inference Using Simulated Data

This section aims to apply the developed algorithms to simulated data sets in a controlled settings. In section 4.4.1, we randomly draw a vector $\theta = (a, b, \sigma_\varepsilon^2)'$ from the

loose prior for a with $\underline{\mu}_a = 1$ and $\underline{V}_a = 100^2$, the elicited priors of σ_ε^2 and b. With vector θ, we simulate a series of data with the SB DGP. Then, using the simulated data series and developed sampling algorithms from section 4.3.1, we carried out the model estimation, the estimation efficiency evaluations, the MCMC convergence diagnostics, and the sampling algorithm efficiency evaluations. The sampling algorithm efficiency is evaluated based on the convergence speed, movement of the Markov chain and autocorrelations among the sequential draws. In section 4.4.2, to compare the Bayesian approach and the classical MLE approach to estimating the SB model, we first arbitrarily choose a vector $\theta = (a, b, \sigma_\varepsilon^2)'$, where the stationary condition is violated. Then, we simulate 100 parallel series' with gradually increasing sample sizes from 20 to 200 and calculated the Root Mean Square Error (RMSE) of the estimates from both the classical MLE approach and the Bayesian approach. The comparison between the classical approach and the Bayesian approach starts with sample sizes as small as 20, because we would like to see if both the classical approach and the Bayesian approach are equally effective in estimating when the data is sparse. In section 4.4.3, using a simulated series from the SB DGP, we carried out the unit root testing using the conventional ADF test and the PP test in the classical framework, and a model comparison in the Bayesian framework using the model marginal likelihood. In section 4.4.3, the Gelfand-Dey algorithm reviewed in section 4.3.2 is applied.

4.4.1 Sampling Efficiency

Following the elicited priors: $p(a) = f_N(1, 10^4)\, 1(|a| < 1)$,
$p(\sigma_\varepsilon^2 \mid a) = f_{\Gamma_\varepsilon}^{-1}(2, 0.1)\, 1(a^2 + \sigma_\varepsilon^2 b^2 < 1)$ and $p(b \mid a, \sigma_\varepsilon^2) = 1(a^2 + \sigma_\varepsilon^2 b^2 < 1)$ [9], we randomly generate a vector of $\theta = (a, b, \sigma_\varepsilon^2)'$ that satisfy the stationary condition, where $a = 0.5094$, $b = 0.2176$ and $\sigma_\varepsilon^2 = 2.0209$.

With the SB DGP,

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. f_N(0, \sigma_\varepsilon^2)$. A series with 200 observations is generated and plotted in Fig (4-1).

Since the series is simulated with parameters generated from the prior and the four moments conditions are easily satisfied, we should expect the high posterior density intervals (HPDI) to cover the 'true' values. Table (4.5) presents the estimated results using the simulated data. With a total of $12,000$ iterations and the first $2,000$ initial draws being discarded, the MCMC converges according to the CD and the NSE

[9] $1(A)$ is the indicator function

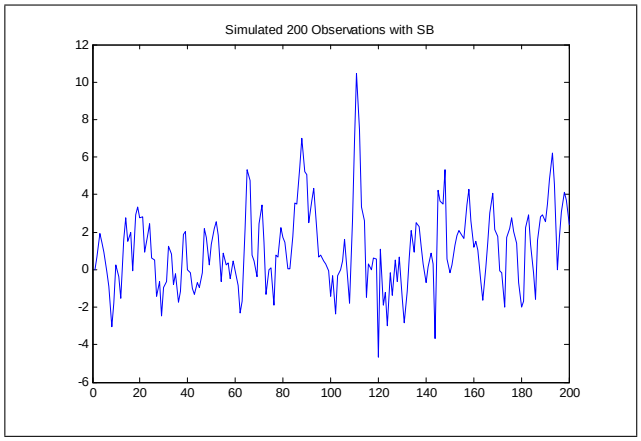


Figure 4-1: Simulated Data with SB DGP Using Elicited Priors

values[10]. From Table (4.5), the true values are within the 95% high posterior density intervals. Therefore, we may conclude that the MCMC converges and the estimation is efficient, and the posterior means are near the set values.

With an efficient sampling algorithm, the initial values should not affect the results after the burn-in. Following the methodologies reviewed in chapter 2.3.3, we evaluate the MCMC efficiency by focusing on the movement of the chain and the correlograms amongst the sequential draws.

Fig (4-2) plots the first 20, 50 and 100 draws from the sampling chains of a, b and σ_ε^2 respectively. Regardless of where the chain starts, the draws of a move towards to the "true" value, 0.5094, quickly. Similarly, although σ_ε^2 starts with value around 3, the Markov chain of σ_ε^2 move swiftly towards to and around the "true" value. We may say that the developed MCMC sampling algorithm in section 4.3.1 is efficient in the sense that the movement in the chain is fast.

Fig (4-3) plots the actual $10,000$ draws of a, b and σ_ε^2 after the first $2,000$ draws were discarded. The correlograms of the draws indicate that the serial correlations of the draws are small after the burn-in. We may also conclude an efficient MCMC sampling algorithm according to the small autocorrelations among the draws.

From Fig(4.a-4.c), we may see that a normal distribution fits the posterior distribution of a well and there is no significant skewness or kurtosis. An Inverse Gamma distribution fits the posterior of σ_ε^2 well. Since the posterior density of b does not have a standard form, from Fig (4.b), the plot of posterior b follows an irregular form, which

[10]The MCMC converges if the CD values are smaller than 1.96 and if the NSE values are near 0 after sufficient iterations.

Table 4.5: Estimates of SB Using Simulated Data

		Prior		*Posterior*					
	$TrueV$	$Mean$	$St.Dev$	$Mean$	$St.Dev$	CD	$NSE_{.15}$	$.025HPDI$	$.975HPDI$
a	0.5094	1	100	0.5803	0.0576	0.3775	0.0005	0.4645	0.6896
b	0.2176	$1(A)$		0.206	0.2279	-0.7001	0.0014	0.1681	0.2441
σ^2_ε	2.0209	$\alpha = 2$	$\beta = 10$	2.1529	0.0022	0.0047	0	1.7600	2.62

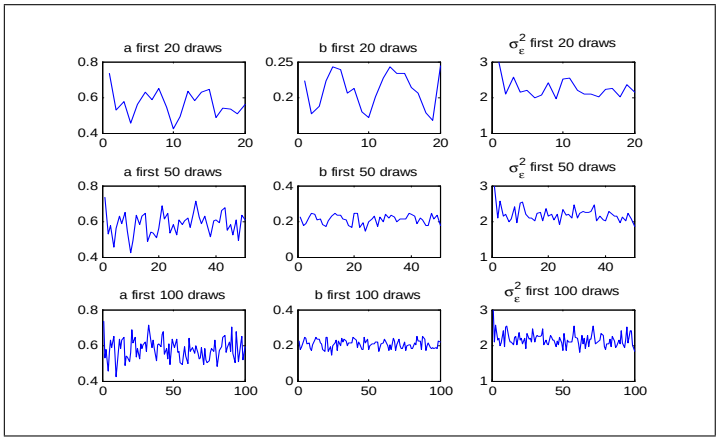


Figure 4-2: Plots of Markov Chains: SB with Simulated Data

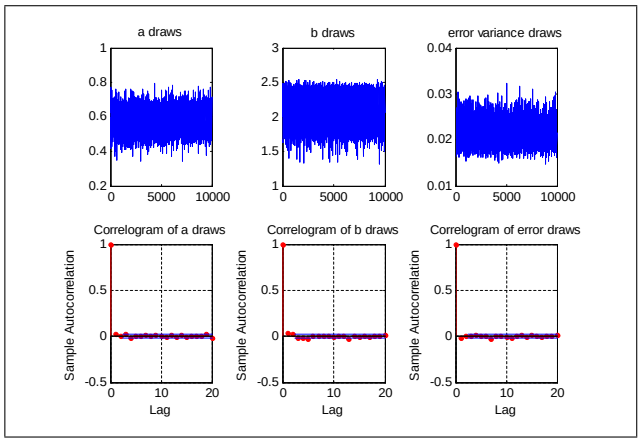


Figure 4-3: Plots of Markov Chains and Corresponding Correlograms: SB with Simulated Data

is multimodal. This shape of b's posterior distribution does not affect our inference or estimations, because we believe that the existence of the multimodal means different values of b's could generate a similar process. From the Bayesian perspective and the inference from the simulated posterior distributions, with a sample size of 200, the Bayesian MCMC estimates do not suffer from huge bias. From the classical perspective, Brunner and Hess (1995) simulate data with $a = 0.5$ and $b = 0.1$ and study the properties of the densities for the maximum likelihood estimates. They found that the densities for a and b are skewed and there is evidence of kurtosis in samples with sizes smaller than 150[11]. However, "as the sample size increase, both the skewness and the kurtosis essentially disappear" (Brunner and Hess, 1995).

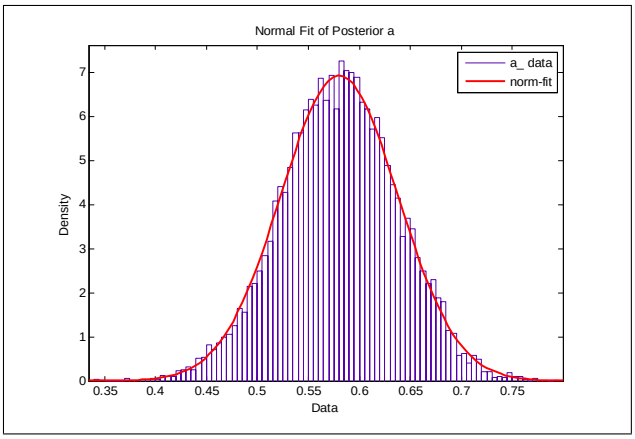


Fig 4.a Normal Fit of Posterior a : SB with Simulated Data

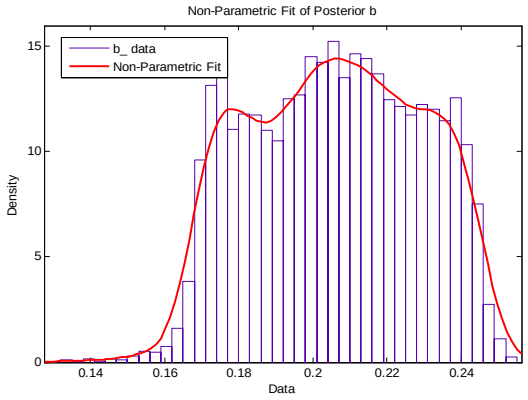


Fig 4.b Non-parametric Fit of Posterior b : SB with Simulated Data

[11]Results can be found in Brunner and Hess (1995), Table 2 on page 672.

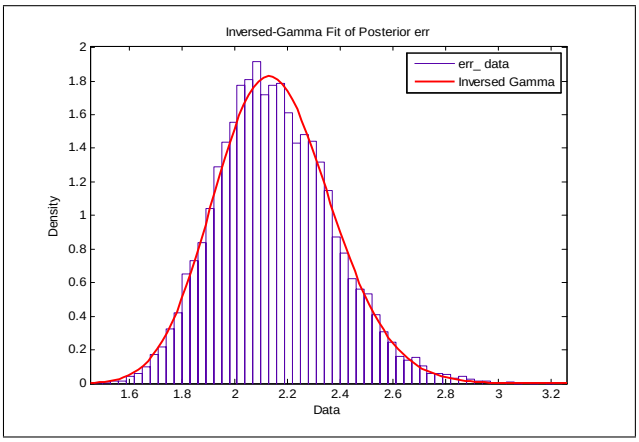


Fig 4.c Inverse Gamma Fit of Posterior σ_{ε}^2 : SB with Simulated Data

This section evaluates the efficiency of the designed sampling algorithm from section 4.3.1 using a simulated data series with a sample size of 200, where the true values easily satisfy the moment restrictions. From the results in Table (4.5) and all the figures above, we may conclude that the designed MCMC sampling algorithm is sufficient to provide efficient samples from the posterior conditionals for Bayesian inference.

4.4.2 Estimations from Classical and Bayesian Methods

This section aims to illustrate how the estimates are affected by the sample sizes in both the classical framework with the MLE and in the Bayesian framework with the MCMC. Also, we would like to demonstrate how the estimates are affected if the elicited priors are hugely distinctive when the Bayesian approach is applied.

To compare the classical MLE, and the Bayesian MCMC estimation method with finite samples, we simulate 100 series with 20 observations, 100 series with 50 observations, 100 series with 100 observations, and another 100 series with 200 observations respectively with the following SB DGP:

$$y_t = (0.9 + 0.0543\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. f_N(0, 0.993)$. The first moment condition is close being violated because the value of a is near 1. Then, all these data series with different sample sizes are estimated using both the classical MLE method and the Bayesian Sampling method.

To illustrate how the elicited priors affect the estimation results, we employed

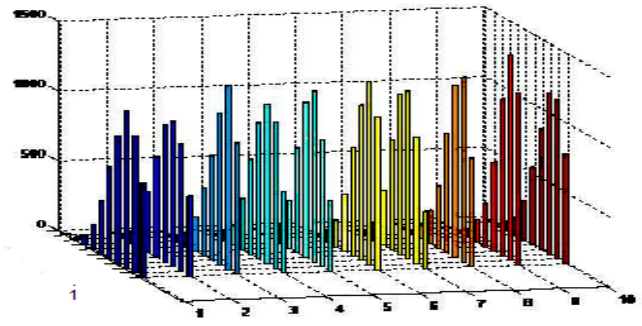


Figure 4-4: Histogram of Posterior a Using 10 Simulated Series

two types of priors for a to estimate when the Bayesian method is employed. One is a relatively loose prior of a, where $p(a) = f_N(1, 100^2)\,1(|a| < 1)$. The other is a very tight prior of a, where $p(a) = f_N(1, 0.1^2)\,1(|a| < 1)$. We also use the prior $p(\sigma_\varepsilon^2 \mid a) = f_{\Gamma_\varepsilon}^{-1}(2, 0.1)\,1(a^2 + \sigma_\varepsilon^2 b^2 < 1)$ and $p(b \mid a, \sigma_\varepsilon^2) = 1(a^2 + \sigma_\varepsilon^2 b^2 < 1)$, which were elicited in section 4.2.2. With a particular sample size, we have the estimated values $\widehat{\beta}_i : i = 1, \cdots, 100$ from the 100 series. Using these 100 estimates, we are able to calculate the Root Mean Square Errors (RMSE) to evaluate the estimates as the following:

$$\widehat{RMSE}\left(\widehat{\beta}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\beta - \widehat{\beta}_i\right)^2}$$

where $N = 100$ and β is the true value. The value of RMSE can help us to evaluate the estimation accuracy based on the estimation bias, and the estimation efficiency based on the variances of the estimates.

We plot the histograms of the posteriors a using the MCMC outputs under the elicited tight priors for 10 randomly simulated series with sample sizes of 20. From Fig (4-4), all the posterior samples of a are within the unity and the posterior means are near 1. From a visual inspection, the highest densities of these 10 posterior distributions are around the 'true' value 0.9, which illustrates that the Bayesian approach is efficient in estimating these 10 random series'.

RMSE results are presented in Table (4.6). Using the loose prior of a, the Bayesian estimation and the Classical estimation perform similarly in terms of estimation accuracy and efficiency. When the sample sizes increase from 20 to 200, the RMSE decreases with both the Bayesian approach and the classical MLE approach.

Comparing the results from an elicited loose prior and an elicited tight prior within the Bayesian framework, RMSE exhibits a big difference between using the loose prior

Table 4.6: Monte Carlo Experiments: RMSE with Classical and Bayesian Approaches to Estimate SB

sample size	Estimationfor100series							
	Classical RMSE		Bayesian RMSE loose prior			tight prior		
				PostMean	Post Mode		Post Mean	Post Mode
$T = 20$	a	0.2211	a	0.2496	0.2078	a	0.0309	0.0512
	b	0.2243	b	0.1547	0.1831	b	0.1723	0.2064
			err	0.6207	0.4619	err	0.2928	0.3569
				Post Mean	Post Mode		Post Mean	Post Mode
$T = 50$	a	0.1057	a	0.1052	0.0996	a	0.0302	0.0394
	b	0.0841	b	0.0831	0.0900	b	0.0834	0.0883
			err	0.2763	0.2286	err	0.1989	0.2161
				Post Mean	Post Mode		Post Mean	Post Mode
$T = 100$	a	0.0660	a	0.0606	0.0526	a	0.0323	0.0333
	b	0.0448	b	0.0491	0.0535	b	0.0484	0.0534
			err	0.1722	0.1521	err	0.1255	0.1345
				Post Mean	Post Mode		Post Mean	Post Mode
$T = 200$	a	0.0429	a	0.0415	0.0402	a	0.0294	0.0321
	b	0.0317	b	0.0389	0.0422	b	0.0312	0.0385
			err	0.1230	0.1229	err	0.0994	0.1017

and the tight prior.

These RMSE results in Table (4.6) show that both the classical approach and the Bayesian approach under a loose prior of a to estimating the SB model perform equally well. The estimation efficiency and accuracy are increased when the sample sizes increase. When a tight prior of a is employed instead of a loose prior, the estimates efficiency are hugely improved with any corresponding sample sizes. We can interpret the results as the following: It is more about the data not being terribly informative and the resulting large variance translates into a "bias" when the parameter space is restricted from above but (effectively) unrestricted from below.

4.4.3 Model Comparison Results from the Marginal Likelihood

From the Bayesian point of view, we may compare the SB model with the RW model even though these two models are not nested. The conventional ADF and PP test may provide misleading results with a seeming-RW process, the SB process.

In this section, a series with 100 observations is generated from the SB DGP where the moment restriction is violated. Firstly, this series is tested for stationarity using

the ADF test and the Phillips-Perron (PP) test in the classical framework. Then, in a Bayesian framework, we fit the data into a SB model and a RW model, then compare these two models via the Bayes Factors. We would like to show that a series generated from the SB DGP behaves like a Random Walk process. This "RW-like" process cannot be distinguished from a RW process using the conventional ADF test or the PP test in the classical framework. However, this "indistinguishable" problem can be compromised in the Bayesian framework, because the model selection procedures in the Bayesian are based on "comparing" rather than "testing".

The SB DGP is as the following:

$$y_t = (0.9689 - 0.0459\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. f_N\left(0, \sigma_\varepsilon^2\right)$, $\sigma_\varepsilon^2 = 10.9513$ is randomly generated from the prior $f_{\Gamma_\varepsilon}^{-1}(2, 0.1)$ Fig (4-5) plots the simulated data series.

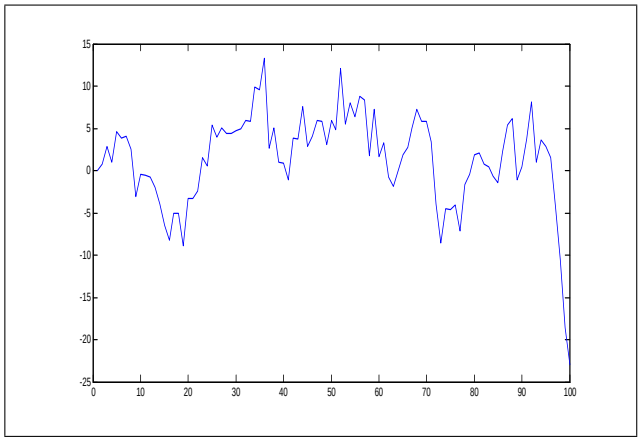


Figure 4-5: Simulated Data: A RW-like Process from the SB DGP

If we pretend without knowing the true DGP, from inspecting the data, a tight prior of a can be employed, recall section 4.2.2 for prior elicitation.

Table (4.7) provides the Unit Root testing results using the conventional ADF test and the PP test. The critical values at 1%, 5% and 10% are also provided. From Table (4.7), the existence of a unit root is not rejected even at the 10% significance level for most of the cases, which indicates the nonstationarity in the data series.

Next, following the Bayesian model comparison procedure, we evaluate the marginal likelihoods of the RW model and the SB model respectively. Then, we calculate the model probabilities via the Bayes Factors.

Table 4.7: ADF and PP Tests for the Unit Root Using the Simulated Series

	ADF			*Phillips − Perron*		
	con	*con&trend*	*none*	*con*	*con&trend*	*none*
test statistic	−1.6425	−1.7437	−1.7803*	−1.6425	−1.7437	−1.7803*
critical value						
1%	−3.4977	−4.0534	−2.5885	−3.4977	−4.0534	−2.5885
5%	−2.8909	−3.4558	−1.9441	−2.8909	−3.4558	−1.9441
10%	−2.5825	−3.1537	−1.6146	−2.5825	−3.1537	−1.6146

* : The null hypothesis of a unit root is rejected.

Fig (4-6) plots the log marginal likelihood of the RW model with the simulated data, where the prior of σ_ε^2 is chosen as an Inverse-Gamma $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$. The log marginal likelihood of the RW model highly depends on the values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$. Although the prior of $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ has been elicited in section 4.2.2 as $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$, we selected a wide range of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ varying from 0.01 to 20 and see how the log marginal likelihood of the RW changes based on different priors, for a robust analysis. From Fig (4-6), the maximum log marginal likelihood is −243.8621 when $\underline{\alpha}_\varepsilon = 20$ and $\underline{\beta}_\varepsilon = 0.01$. The prior density of $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(20, 0.01)$ is plotted in Fig (4-7.a). We may see that the highest density under this prior is around 5 and the probabilities of σ_ε^2 being a value larger than 7 or smaller than 3 are very low. Therefore, the prior $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(20, 0.01)$ will be very subjective if it had been chosen to estimate the RW model. In other words, we would not choose prior $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(20, 0.01)$ for the RW model in the first place to estimate.

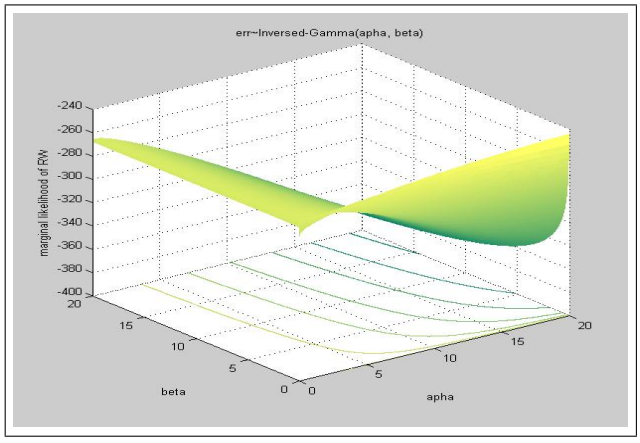


Figure 4-6: log Marginal Likelihood of RW with Simulated Data

To compare the RW model and the SB model, we select the same prior for the

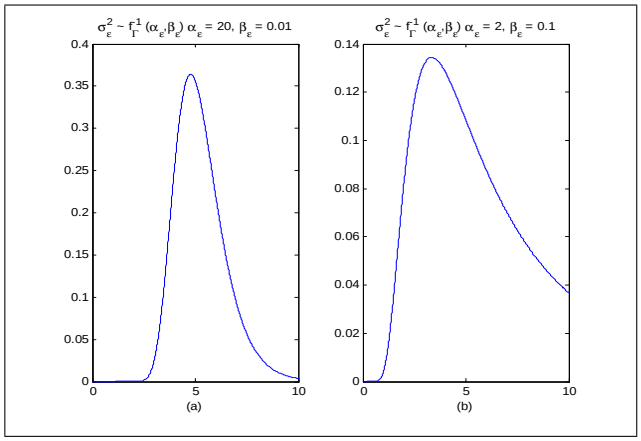


Figure 4-7: Prior Densities for σ_ε^2 with Different Values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$

Table 4.8: log Marginal Likelihood of SB with Gelfand-Dey Method

Marginal Likelihood $\underline{\alpha}_\varepsilon$ fixed as 2			
$\underline{\beta}_\varepsilon$	$p = 0.01$	$p = 0.02$	$p = 0.03$
200	-273.7574	-272.9708	-272.5025
20	-270.5681	-270.4196	-270.3697
2	-265.3359	-265.0199	-264.8950
0.5	-262.2970	-261.6896	-261.3697
0.3	-261.3569	-260.6002	-260.1577
0.1	-260.9438	-260.1944	-259.7419

common parameter $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$. This prior represent a belief that the σ_ε^2 could be a value from a wide range because the distribution has high densities towards to the right tail (Fig 4-7.b plots the prior density of $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(2, 0.1)$). Also, this prior of the RW model has been elicited in section 4.2.2 based on simulations. With this elicited prior, the log marginal likelihood of RW is -267.5175.

Table (4.8) reports the log marginal likelihood results with SB model using the simulated data, where the Gelfand-Dey method is applied. Similar to the RW model, although $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ has been elicited in section 4.2.2 where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$ for the SB model, to illustrate how the choices of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ affect the log marginal likelihood with the SB model, we fixed $\underline{\alpha}_\varepsilon$ as 2 but allow $\underline{\beta}_\varepsilon$ to vary at different values from 20 to 0.1 for a robust analysis. Then we calculate the log marginal likelihood of the SB under different selections of $\underline{\beta}_\varepsilon$.

In Fig (4-8), we also plot the prior densities with a range of $\underline{\beta}_\varepsilon$ when $\underline{\alpha}_\varepsilon$ is fixed as

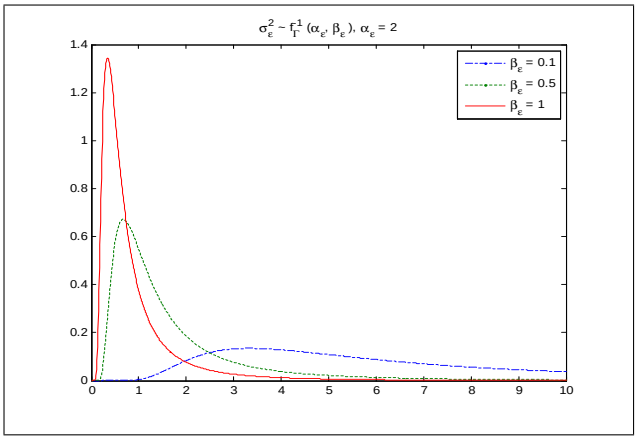


Figure 4-8: Prior Densities of σ_ε^2 with $\underline{\alpha}_\varepsilon = 2$ and Various $\underline{\beta}_\varepsilon$

2. From Fig (4-8), the scale of σ_ε^2 increases when $\underline{\beta}_\varepsilon$ gets smaller. If the priors of σ_ε^2 is chosen with a large $\underline{\beta}_\varepsilon$, clearly the prior will be so tight that it could not properly represent our belief in σ_ε^2.

From Table (4.8), the log marginal likelihood is higher with smaller $\underline{\beta}_\varepsilon$. Using a common prior $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(2, 0.1)$ for the SB model and the RW model, the log marginal likelihood of the SB model is -259.7419. With a log marginal likelihood from the RW model at -267.5175, the log Bayes Factor will then be $\log BF_{RW:SB} = -7.7756$, which provides strong evidence of supporting the SB model based on the data information.

4.5 Empirical Applications: a Study of the UK Inflation Rates

In this section, we apply the posterior simulator developed in section 4.3.1 and the Gelfand-Dey method reviewed in section 4.3.2 with the real life macroeconomic data. The data we applied is the quarterly UK headline inflation rate from 1957 quarter one to 2007 quarter one. This empirical application has two purposes: One is to fit the SB model with the data and compare the differences of the estimates between the Bayesian MCMC method and the classical MLE method. The other is to evaluate the uncertainties of the nonstationarity in the data series by comparing the nonlinear SB model with the linear RW model using the log marginal likelihoods.

As motivation, we briefly review the literature concerning the UK inflation rates in time series modelling. Fig (4-9) plots the quarterly "headline" UK inflation rates,

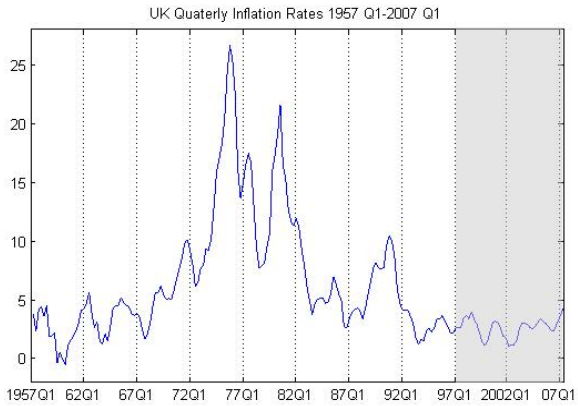


Figure 4-9: Data Plot: UK Quarterly Inflation Rates (1957Q1-2007Q1)

which is the annual percentage change in the seasonally unadjusted Retail Price Index (RPI). The RPI is an all-inclusive RPI, which also includes the mortgage payments[12].

As we mentioned in the introduction, inspired by Watson and Stock (2007), we apply the SB model, which holds a time varying parameter feature, with the inflation rates. To motivate our studies in the UK's experiences, we firstly provide a brief review of the important historical events and monetary policies in the UK after 1950s. Then, we review the findings in the literature regarding the UK inflation rates modelling. For a thorough review of the economy and political history in the UK before the 1980's, refer to Caves (1980).

King (1997) summarizes the univariate inflation in the UK

> Over the three hundred years since the Bank of England was founded in 1694, the average inflation in Britain has been 1.4%. But in the period since the second world war, inflation has averaged 6% and between 1965 and 1980 it averaged no less than 10.3%. Since 1945 prices have risen more than twenty-fold. Creeping inflation in the 1950s and early 1960s led to rapid inflation in the 1970s, reaching a peak of 27% in August 1975, before a gradual disinflation during the 1980s and 1990s.

The sample period we applied in this chapter is from 1957-2007. Within this period, the Bretton Wood system broke down (1970-1971), and Sterling departed from the Exchange Rate Mechanism (1992). Also, there were two oil crises (1973-1975 and

[12] At the moment, in the UK under the inflation targeting policy, the target range for inflation actually is defined in terms of the RPIX, which is the RPI excluding the mortage payments.

1979-1980). The UK experienced high rates of inflation with a low output between the early 1960's and the mid 1980's. This phenomenon is also known as the "Great Inflation" or the "Great Stagflation". After 2003, the inflation rate in the UK is low and stable while the economy is growing at a satisfactory rate. Over the last four decades, the UK not only had an anguished history of disinflation, but also the UK is one of the pioneering countries that has applied inflation targeting and made the central bank independent[13]. Moreover, the UK has richer data resources than many other countries. Therefore, it is sensible that learning from the historical experiences of the UK would help policy makers avoid similar episodes in the future. Next, we summarize the findings based on the UK's experience in the literature.

Hendry (2001) models the long run UK inflation rate from 1875 to 1991. He proposes an "eclectic model", in which a great number of economic variables affect inflation jointly.

Using purely statistical models, the underlying process of the UK inflation is well documented. Kapetanios (2004) proposes the factor model to forecast the core inflation rate based on the evidence of using a large disaggregated price index data set. Henry and Shields (2004) modelled the UK inflation rates as a two-regime threshold unit root process. They found that shocks to inflation are highly persistent in one regime but have finite lives in the other. Relevant literature also includes the analysis of the variations of the inflation. Baillie et al. (1996) applied a so-called autoregressive fractionally integrated moving average-generalized auto-regressive conditional heteroscedasticity (ARIFMA-GARCH) model to fit the monthly CPI inflation and the inflation uncertainties. They found that in the UK, there is strong evidence of a joint feedback between the conditional mean and variance of inflation, which support the Friedman (1977) hypothesis. Kontonikas (2002) also applied the GARCH model to evaluate the inflation and inflation uncertainties for a long run UK data. His results indicate a positive relationship between past inflation and current uncertainty.

[13] Actually, the new framework for monetary policy commenced in 1992, following the UK leaving the Exchange Rate Mechanism. This new framework includes an explicit target of inflation and increasing transparency in the Bank of England. However, the Bank of England was made independent only after the Bank of England Act (1998) in a sense that the Bank of England has the power to set interest rates. See the following link for details. http://en.wikipedia.org/wiki/Bank_of_England

4.5.1 Estimation with a SB Model Specification

After a short review regarding the UK inflation rate, we fit the data series with a sample size of 201 into the simplest first order SB model:

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. f_N(0, \sigma_\varepsilon^2)$. Table (4.9) provides the Bayesian MCMC estimation results from $12,000$ iterations with the first $2,000$ draws discarded with a loose prior of a. The chain converges according to the CD value and NSE value. Table (4.10) provides the estimation results from the classical MLE method.

From Tables (4.9) and Table (4.10), the estimates using the Bayesian MCMC method under a loose prior of a do not deviate a lot from the results using the classical method. Similar results have been shown under the controlled setting using the simulated series (see section 4.4.2).

Fig (4-10) plots the actual draws from the chain, the correlograms and the histograms of a, b, and σ_ε^2. From Fig (4-10), posterior distribution of a can be fitted with a truncated normal distribution. The posterior mean of a is below but very close to unity. The posterior of b shows that b has large probabilities of being around 0.04. The posterior of σ_ε^2 can be fitted with an Inverse-Gamma distribution, which has high densities around 1.5. According to the correlograms, we consider the proposed MCMC to be efficient that approximates the target distribution well and extra iterations will not be necessary.

Both results from the Bayesian MCMC approach and the classical MLE approach indicate that a is near to unity and a bilinearity term b exists. This non-zero b would indicate that the inflation series has a very long memory such that all the past shocks have impacts on the current inflation status.

4.5.2 Model Selections Using the SB Model and the RW model

This section will demonstrate that, in a real settings, the classical method of hypothesis "testing" and the Bayesian method of model "comparing" may arrive at different answers. Firstly, we carried out the conventional ADF test and the PP test with the UK inflation series. Then we fit the data with a RW model and a first order SB model. Finally, we calculate the log marginal likelihoods and compare these two models in the Bayesian context.

Table 4.9: Bayesian Approach: Estimates of SB with Quarterly UK Inflation Rates

	Prior		*Posterior*						
	Mean	*St.Dev*	*Mean*	*St.Dev*	*CD*	$NSE_{.15}$	*Median*	95%*Posterior Band*	
a	1	100	0.9769	0.0133	−0.0458	0.0001	0.9783	0.9472	0.9974
b	$1(A)$		0.0395	0.0078	−0.0898	0.0064	0.0403	0.0230	0.0513
σ_{ε}^2	$\underline{\alpha}_{\varepsilon}=2$	$\underline{\beta}_{\varepsilon}=0.1$	1.4217	0.1467	1.7798	0	1.4092	1.1635	1.7346

† :where $1\left(|b|<\sqrt{(1-a^2)\,h_{\varepsilon}}\right)$ is the indicator function for

Table 4.10: Classical Approach: Estimates of SB with Quarterly UK Inflation Rates

	ClassicalEstimates			
parameter	*Mean*	*st.error*	$t-value$	$Max.\lg likelihood$
a	0.9808	0.0163	60.0119	-309.7733
b	0.0483	0.002	22.7849	s.e. of err: 1.13

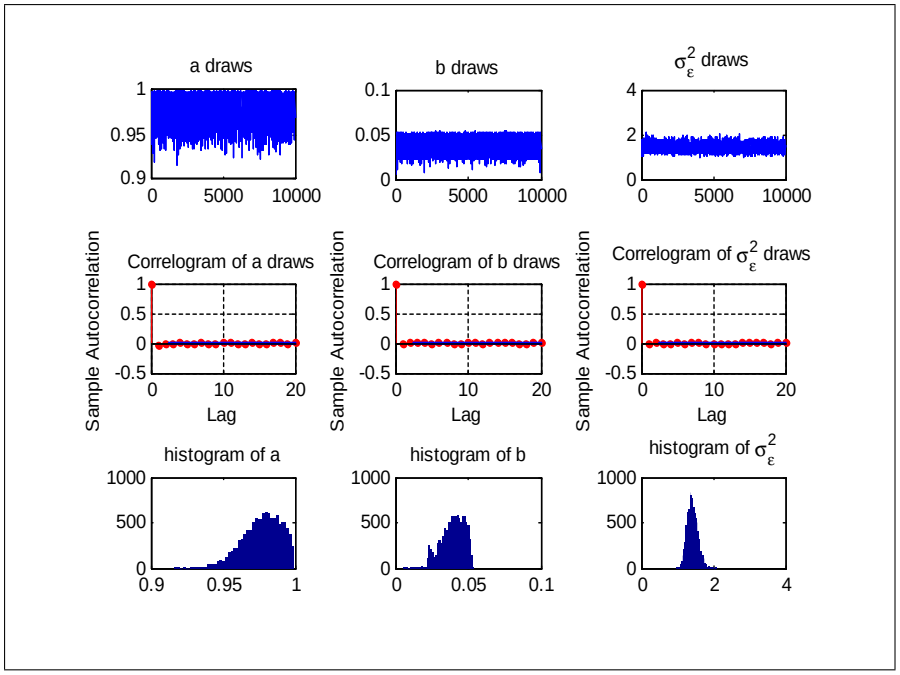


Figure 4-10: Chains Plots of a, b, and σ_{ε}^2, Corresponding Correlograms and Histograms

Table 4.11: ADF and PP Tests with the Quarterly UK Inflation Rates

	ADF			Phillips-Perron		
	cons	*cons&trend*	*none*	*cons*	*cons&trend*	*none*
test statistic	-1.8336	-1.9250	-1.2188	-2.4979	-2.5939	-1.5403
critical value						
1%	-3.4416	-3.9742	-2.5690	-3.4413	-3.9738	-2.5689
5%	-2.8664	-3.4177	-1.9414	-2.8663	-3.4175	-1.9414
10%	-2.5694	-3.1313	-1.6163	-2.5693	-3.1312	-1.6163

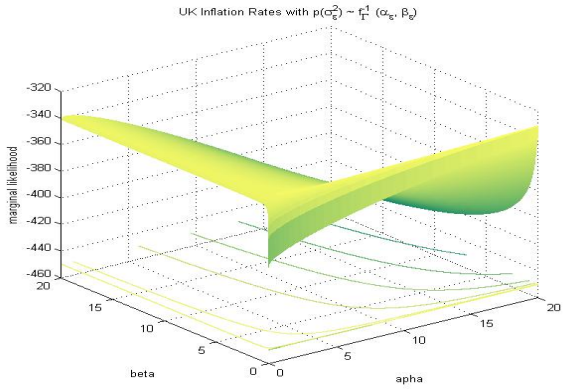


Figure 4-11: log Marginal Likelihood of RW with UK Inflation Rates

Table (4.11) provides the unit root test results using the conventional ADF test and the PP test. The null hypothesis of a unit root in the data series is neither rejected by the conventional ADF test nor the PP test. From the classical perspective, nonstationarity exists in the underlying process of the data series.

Following the exact same model comparison procedure as that in the controlled settings (see section 4.4.3), we calculate the log marginal likelihood by fitting the RW model and the SB model with the data series. Fig (4-11) plots the log marginal likelihood of the RW model with a prior of $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, where both $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ vary from 0.01 to 20.

From Fig (4-11), the log marginal likelihood is maximized at -336.4855 when $\underline{\alpha}_\varepsilon = 20$ and $\underline{\beta}_\varepsilon = 0.03$. A prior of $\sigma_\varepsilon^2 \sim f_\Gamma^{-1}(20, 0.03)$ indicates a high density around 1.5, as plotted in Fig (4-13). With elicited prior in section 4.2.2, where $\underline{\alpha}_\varepsilon = 2$ and $\underline{\beta}_\varepsilon = 0.1$, the log marginal likelihood will be -339.2917. To illustrate how the change of $\underline{\alpha}_\varepsilon$ affect the log marginal likelihood with the RW model, we plot the prior densities with $\underline{\beta}_\varepsilon$ fixed as 0.1 and $\underline{\alpha}_\varepsilon$ varies from 1 to 20 in Fig (4-12).

From Fig (4-12 and 4-13), the value of $\underline{\alpha}_\varepsilon$ affects the scale of the prior density. Large values of $\underline{\alpha}_\varepsilon$, such as 20, limits the probability of σ_ε^2 being larger than 1.5. Therefore, we have to select a prior $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, which represents a prior belief that σ_ε^2 could be a value larger than 5. Clearly, $\underline{\alpha}_\varepsilon$ can be chosen as 1 or 2 to satisfy our requirement. Then, we provide the log marginal likelihood with the RW model under different combinations of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ in Table (4.12).

Using the Gelfand-Dey method with a prior of $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}(2, 0.1)$, using the truncated tails at 1%, 2% and 3%, the log Marginal Likelihood is provided in Table (4.13). From the log Bayes factors, the SB model is more favoured over the RW model. Under

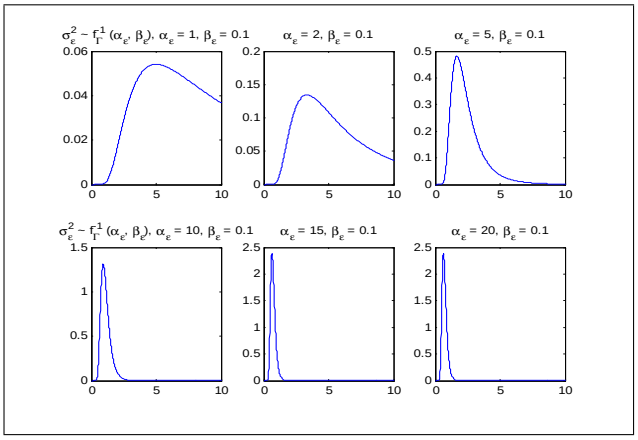


Figure 4-12: Density Plots of σ_{ε}^2 with $\underline{\beta}_{\varepsilon} = 0.1$ and Varying $\underline{\alpha}_{\varepsilon}$

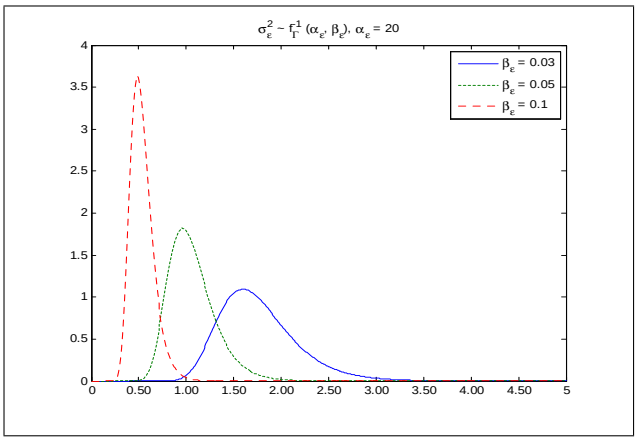


Figure 4-13: Density Plots of σ_{ε}^2 with $\underline{\alpha}_{\varepsilon} = 20$ and Varying $\underline{\beta}_{\varepsilon}$

Table 4.12: log Marginal Likelihood with RW under Different Priors: Quaterly UK Inflation Rates

	α_{ε}= 10				
β_{ε}	20	15	10	5	1
lg $MLRW$	−382.2503	−379.3839	−375.3501	−368.4812	−352.8854
	β_{ε}= 0.5				
α_{ε}	10	8	6	4	2
lg $MLRW$	−359.5055	−353.3608	−347.7455	−342.8623	−339.1732

Table 4.13: log Marginal Likelihood of SB and lgBF Compared with RW: UK Qaurterly Inflation Rates

pp.	0.01	0.02	0.03
lg $MLSB$	-257.5976	-256.4027	-255.4636
Comparing with RW at the lgMLmax -339.2917			
logBF	81.694	82.889	83.828

the real settings with an application of the quarterly UK inflation rates, the conventional ADF test and the PP test fail to reject the null hypothesis of a unit root in the data series. However, the Bayes factors could quantify the uncertainties via comparing models' probabilities of the entertained models.

4.6 Conclusion

In this chapter, we focus on a bilinear model, which can represent most well-behaved nonlinearities. A simple first order bilinear model, the SB model is investigated in the Bayesian context. As the SB model allows for the shocks in the system to affect the persistence in the underlying process, we agree that the SB model could be applied with the macroeconomic series, such as the inflation rates, to model their underlying processes. Because the data series generated from the SB DGP are RW-like processes, the conventional ADF and the PP tests may not be able to distinguish the SB process from a RW process. Therefore, testing results from the ADF and the PP test could potentially provide very misleading inference, and the misinterpretation of the bilinear term b may induce severe distorted forecasts.

The contributions of this chapter are as follows:

First, motivated by Brunner and Hess (1995) regarding the potential problems in estimating the bilinear model with the classical MLE, we developed an efficient MCMC sampling algorithm to estimate the Stationary Bilinear model in a Bayesian context. When the moment restrictions are volatile, such as when a is near the unity, the classical MLE using the NR-SA algorithm performs equally well as the Bayesian MCMC method under a loose prior of a. When the sample size increases from 20 to 200, the estimation efficiencies from both approaches are increased. Choosing a tight prior of a using the Bayesian MCMC to estimate can gain huge improvements in the estimation accuracy with all corresponding sample sizes, when the stationary condition is violated. When the data sample is under 100, if the series is a RW-like process from a simple visual inspection, we could apply the SB model with a tight

prior to model the series' underlying process. Practitioners should pay more attention when eliciting the priors. In practice, using simulations may shed light on the prior elicitations.

Second, the classical ADF test and the PP test must be applied with caution when the process is a RW-like process, e.g. the SB process, because these classical tests tend to mistake the SB process with a nonstationary unit root process. As a consequence, the misinterpreting in the bilinear term b may lead to incorrect inference about the conditional heteroscedasticity and distorted forecasts. In this chapter, we employed the Gelfand-Dey method to evaluating the log marginal likelihood of the SB model, which enables the practitioners to compare amongst the RW model, the SB model or any other entertained models where the marginal likelihood calculation is available, conditional on the available data information.

Finally, we revisit the quarterly UK headline inflation rates via fitting the SB model and the RW model with the data. Estimation results show that the UK inflation rate is highly persistent, and the past shocks play an important role in forming the current inflation because of a non-zero bilinear term. While the unit root is not rejected with the conventional ADF and the PP test in the classical framework, the SB model receives a higher model probability when compared with the RW model in the Bayesian framework.

For future studies, we may extend this simple first order SB model in Equation (4.1) by incorporating the structure breaks and changing volatilities. The new extended model will be as follows:

$$y_t = (a_k + b_k \varepsilon_{t-1})\, y_{t-1} + \varepsilon_t \tag{4.16}$$

where $\varepsilon_t \sim i.i.d. f_N(0, \sigma_k^2)$ and $k = 1, 2$. Since the forecast of y_t from the extended model will be $E(y_t) = (a_k + b_k \varepsilon_{t-1})\, y_{t-1}$, if there is a break, or changes in the disturbance property, the extended model may have more resilience and be more adjustable to accommodating breaks.

Appendix 4.A Recursive Derivation of SB Model's Likelihood

The first order SB model is

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $t = 2, \cdots, n$, and $\varepsilon_1 = 0$. If $y = (y_1, \cdots, y_n)'$ and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$

$$y_2 = (a + b\varepsilon_1) y_1 + \varepsilon_2$$

when $t = n$, $n \geq 3$

$$y_n = a y_{n-1} + b y_{n-1} \varepsilon_{n-1} + \varepsilon_n$$

then we have

$$\varepsilon_{n-1} = y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2}$$

and

$$\varepsilon_{n-2} = y_{n-2} - a y_{n-3} - b y_{n-3} \varepsilon_{n-3}$$

If we take the substitute equations recursively

$$\begin{aligned}
y_n &= a y_{n-1} + b y_{n-1} \left(y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2} \right) + \varepsilon_n \\
&= a y_{n-1} + b y_{n-1} y_{n-1} - ab y_{n-1} y_{n-2} - b^2 y_{n-1} y_{n-2} \varepsilon_{n-2} + \varepsilon_n \\
&= a y_{n-1} + b y_{n-1} y_{n-1} - ab y_{n-1} y_{n-2} - b^2 y_{n-1} y_{n-2} \left(y_{n-2} - a y_{n-3} - b y_{n-3} \varepsilon_{n-3} \right) + \varepsilon_n \\
&= a y_{n-1} + b y_{n-1} y_{n-1} - ab y_{n-1} y_{n-2} - b^2 y_{n-1} y_{n-2} y_{n-2} + ab^2 y_{n-1} y_{n-2} y_{n-3} \\
&\quad + b^3 y_{n-1} y_{n-2} y_{n-3} \varepsilon_{n-3} + \varepsilon_n
\end{aligned} \tag{4.17}$$

where the last expanded form with ε_2 is

$$\begin{aligned}
& (-1)^{n-1} \left[b^{n-2} y_{n-1} y_{n-2} \cdots y_2 \varepsilon_2 \right] \\
= \; & (-1)^{n-1} \left[b^{n-2} y_{n-1} y_{n-2} \cdots y_2 \left(y_2 - a y_1 - b \varepsilon_1 y_1 \right) \right] \\
= \; & (-1)^{n-1} b^{n-2} y_{n-1} \cdots y_2 y_2 + a \left(-b \right)^{n-2} y_{n-1} y_{n-2} \cdots y_1 - (-1)^n b^{n-1} y_{n-1} y_{n-2} \cdots y_1 \varepsilon_1
\end{aligned}$$

Since $\varepsilon_1 = 0$, Equation (4.17) can be generalized for $t = 3, \cdots, n$ as

$$y_t = a \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right] + \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right] + \varepsilon_t$$

If we denote

$$f_1 (t, b) = \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right]$$

$$f_2 (t, b) = \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right]$$

thus $t = 3, \cdots, n$

$$y_t = af_1(t,b) + f_2(t,b) + \varepsilon_t$$

and for $t = 2,\ f_2(t,b) = 0$

$$y_t = af_1(t,b) + \varepsilon_t$$

Hence, we have the likelihood function as the following:

$$p(y|a,b,h_\varepsilon) = \frac{h_\varepsilon^{\frac{N-1}{2}}}{(2\pi)^{\frac{N-1}{2}}} \exp\left\{-\frac{h_\varepsilon}{2}\sum_{t=2}^{N}\left[y_t - af_1(t,b) - f_2(t,b)\right]^2\right\}$$

Derivation of Posterior Conditionals

According to the likelihood and specified prior densities (see details in section 4.2.1), the full joint posterior can be derived as the following:

$$\begin{aligned}
p(a,b,h_\varepsilon|y) &\propto p(y|a,b,h_\varepsilon)\,p(a,b,h_\varepsilon) \\
&\propto \frac{h_\varepsilon^{\frac{N-1}{2}}}{(2\pi)^{\frac{N-1}{2}}} \exp\left\{-\frac{h_\varepsilon}{2}\sum_{t=2}^{N}\left[y_t - af_1(t,b) - f_2(t,b)\right]^2\right\} \\
&\quad \cdot \frac{1}{\Pr(1\,||a|<1|)\sqrt{2\pi\underline{V}_a}} \exp\left\{-\frac{\left(a-\underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} \cdot 1(|a|<1) \\
&\quad \cdot \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma(\underline{\alpha}_\varepsilon)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) \cdot 2\sqrt{(1-a^2)\,h_\varepsilon} \\
&\quad \cdot \frac{1}{2\sqrt{(1-a^2)\,h_\varepsilon}} \cdot 1\left[|b| < \sqrt{(1-a^2)\,h_\varepsilon}\right] \qquad (4.18)
\end{aligned}$$

1. **Derivation of the Posterior Conditionals of** a

$$\begin{aligned}
p(a|y,b,h_\varepsilon) &\propto \exp\left\{-\frac{h_\varepsilon}{2}\sum_{t=2}^{N}\left[y_t - af_1(t,b) - f_2(t,b)\right]^2\right\} \\
&\quad \cdot \exp\left\{-\frac{\left(a-\underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} 1\left[|b| < \sqrt{(1-a^2)\,h_\varepsilon}\right] \qquad (4.19)
\end{aligned}$$

then

$$p\left(a|y,b,\sigma_\varepsilon^2\right) \propto 1\left[|a|<\sqrt{1-b^2\sigma_\varepsilon^2}\right] \cdot \exp\left\{-\frac{1}{2}\left[\underbrace{\frac{\sum_{t=2}^{N}\left[y_t-af_1\left(t,b\right)-f_2(t,b)\right]^2}{\sigma_\varepsilon^2}+\frac{\left(a-\underline{\mu}_a\right)^2}{\underline{V}_a}}_{A}\right]\right\}$$

where

$$\begin{aligned} A &= a^2\cdot\frac{\sum_{t=2}^{N}f_1^2\left(t-1,b\right)}{\sigma_\varepsilon^2}-2a\cdot\frac{\sum_{t=2}^{N}f_1\left(t,b\right)\left[y_t-f_2\left(t,b\right)\right]}{\sigma_\varepsilon^2}+\frac{a^2}{\underline{V}_a}-\frac{2a\underline{\mu}_a}{\underline{V}_a} \\ &= a^2\left\{\frac{\sum_{t=2}^{N}f_1^2\left(t,b\right)}{\sigma_\varepsilon^2}+\frac{1}{\underline{V}_a}\right\}-2a\left\{\frac{\sum_{t=2}^{N}f_1\left(t,b\right)\left[y_t-f_2\left(t,b\right)\right]}{\sigma_\varepsilon^2}+\frac{\underline{\mu}_a}{\underline{V}_a}\right\} \end{aligned}$$

Therefore, the posterior variance $\overline{V}_a$ of a

$$\overline{V}_a=\left[\frac{\sum_{t=2}^{N}f_1^2\left(t,b\right)}{\sigma_\varepsilon^2}+\frac{1}{\underline{V}_a}\right]^{-1}$$

and the posterior mean $\overline{\mu}_a$ of a is

$$\overline{\mu}_a=\overline{V}_a\cdot\left[\frac{\sum_{t=2}^{N}f_1\left(t,b\right)\left[y_t-f_2\left(t,b\right)\right]}{\sigma_\varepsilon^2}+\frac{\underline{\mu}_a}{\underline{V}_a}\right]$$

2. **Derivation of the Posterior Conditionals for** h_ε

According to Equation (4.18)

$$
\begin{aligned}
p\left(h_{\varepsilon}|y,a,b\right) &\propto \frac{h_{\varepsilon}^{\frac{N-1}{2}}}{(2\pi)^{\frac{N-1}{2}}}\exp\left\{-\frac{h_{\varepsilon}}{2}\sum_{t=2}^{N}\left[y_t-af_1\left(t,b\right)-f_2(t,b)\right]^2\right\} \\
&\quad \cdot\frac{1}{\underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}}\Gamma\left(\underline{\alpha}_{\varepsilon}\right)}h_{\varepsilon}^{\underline{\alpha}_{\varepsilon}-1}\exp\left(-\frac{h_{\varepsilon}}{\underline{\beta}_{\varepsilon}}\right)\cdot 1\left[|b|<\sqrt{\left(1-a^2\right)h_{\varepsilon}}\right] \\
&\propto 1\left[h_{\varepsilon}>\frac{b^2}{1-a^2}\right]\cdot h_{\varepsilon}^{\underline{\alpha}_{\varepsilon}+\frac{N-1}{2}-1} \\
&\quad \cdot\exp\left\{-h_{\varepsilon}\left[\frac{1}{\underline{\beta}_{\varepsilon}}+\frac{1}{2}\sum_{t=2}^{N}\left[y_t-af_1\left(t,b\right)-f_2(t,b)\right]^2\right]\right\}
\end{aligned}
$$

Thus, we need random draws of h_{ε} from a truncated Gamma distribution with $f_{\Gamma_{\varepsilon}}\left(\overline{\alpha}_{\varepsilon},\overline{\beta}_{\varepsilon}\right)\cdot 1\left[h_{\varepsilon}>\frac{b^2}{1-a^2}\right]$, where

$$
\overline{\beta}_{\varepsilon}=\left[\frac{1}{\underline{\beta}_{\varepsilon}}+\frac{1}{2}\sum_{t=2}^{N}\left[y_t-af_1\left(t,b\right)-f_2(t,b)\right]^2\right]^{-1}
$$

and

$$
\overline{\alpha}_{\varepsilon}=\underline{\alpha}_{\varepsilon}+\frac{N-1}{2}
$$

Appendix 4.B Joint Prior density at Posterior Draws

Using the Gelfand-Dey method, joint prior densities $p\left(a,b,h_{\varepsilon}\right)$ are evaluated at posterior draws $p\left(\mathbf{a},\mathbf{b},\boldsymbol{\sigma}_{\varepsilon}^2\right)$ at $a^{(g)*},b^{(g)*},h_{\varepsilon}^{(g)*}$ where $g=1,...G$. Given the joint prior densities, the joint prior evaluated at each posterior draw can be simplified as the following:

$$
\begin{aligned}
& p\left(a^{(g)*}, b^{(g)*}, h_{\varepsilon}^{(g)*}\right) \\
= \; & p\left(b^{(g)*} \mid h_{\varepsilon}^{(g)*}, a^{(g)*}\right) p\left(h_{\varepsilon}^{(g)*} \mid a^{(g)*}\right) p\left(a^{(g)*}\right) / C_t \\
= \; & \frac{1}{\sqrt{2\pi \underline{V}_a}} \exp\left\{-\frac{\left(a^{(g)*}-\underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} \cdot 1\left(\left|a^{(g)*}\right|<1\right) \cdot \frac{1}{\Pr\left(\left|N(\underline{\mu}_a, \underline{V}_a)\right|<1\right)} \\
& \cdot C_{\Gamma_\varepsilon}^{-1} h_{\varepsilon}^{\frac{\underline{\upsilon}_\varepsilon-2}{2}} \exp\left(-\frac{h_{\varepsilon}^{(g)*}\underline{\upsilon}_\varepsilon \underline{s}_\varepsilon^2}{2}\right) \cdot 2\sqrt{\left(1-a^{(g)*2}\right) h_{\varepsilon}^{(g)*}} \\
& \cdot \frac{1}{2\sqrt{\left(1-a^{(g)*2}\right) h_{\varepsilon}^{(g)*}}} \cdot 1\left[\left|b^{(g)*}\right| < \sqrt{\left(1-a^{(g)*2}\right) h_{\varepsilon}^{(g)*}}\right] \\
= \; & \frac{1}{\sqrt{2\pi \underline{V}_a}} \cdot C_{\Gamma_\varepsilon}^{-1} \exp\left\{-\frac{\left(a^{(g)*}-\underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} \cdot \frac{1}{\Pr\left(\left|N(\underline{\mu}_a, \underline{V}_a)\right|<1\right)} \\
& \cdot h_{\varepsilon}^{\frac{\underline{\upsilon}_\varepsilon-2}{2}} \exp\left(-\frac{h_{\varepsilon}^{(g)*}\underline{\upsilon}_\varepsilon \underline{s}_\varepsilon^2}{2}\right) \cdot 1\left[\left|b^{(g)*}\right| < \sqrt{\left(1-a^{(g)*2}\right) h_{\varepsilon}^{(g)*}}\right]
\end{aligned}
$$

where

$$\frac{\underline{\upsilon}_\varepsilon}{2} = \underline{\alpha}_\varepsilon$$

and

$$\frac{\underline{\upsilon}_\varepsilon \underline{s}_\varepsilon^2}{2} = \frac{1}{\underline{\beta}_\varepsilon}$$

Appendix 4.C Working with Transformed Data

If the data is transformed, the prior selections should be modified. The original model is as the following:

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$.

Because the data we are analyzing is the UK's inflation rates, the data measured on percentage should be at most 0.3, which indicates the monthly inflation rates (price level comparing to the same month in the last year) is not exceeding 30%. Then, we transform the data by dividing the data with s, where $s = 100$:

$$\frac{y_t}{s} = (a + b\varepsilon_{t-1})\, \frac{y_{t-1}}{s} + \frac{\varepsilon_t}{s}$$

$$y_t^* = \left(a + b^* \times \varepsilon_{t-1}^*\right) y_{t-1}^* + \varepsilon_t^*$$

where

$$y_t^* = \frac{y_t}{s}, y_{t-1}^* = \frac{y_{t-1}}{s}, \varepsilon_t^* = \frac{\varepsilon_t}{s}$$

$$\varepsilon_{t-1}^* = \frac{\varepsilon_{t-1}}{s}, b^* = sb$$

This indicates that $\varepsilon_t^* \sim (0, \sigma_\varepsilon^{2*})$, where $\sigma_\varepsilon^{2*} = \frac{\sigma_\varepsilon^2}{s^2}$.Since the stationary conditions still hold with the transformed data, we have $a^2 + b^{2*}\sigma_\varepsilon^{2*} < 1$. If we denote $S = s^2$, then $\sigma_\varepsilon^{2*} = g\left(\sigma_\varepsilon^2\right) = \sigma_\varepsilon^2/S$ and the following

$$f\left(\sigma_\varepsilon^2\right) = \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)\sigma_\varepsilon^{2(\underline{\alpha}_\varepsilon+1)}} \exp\left(-\frac{1}{\underline{\beta}_\varepsilon\sigma_\varepsilon^2}\right)$$

$$\begin{aligned} f\left(\sigma_\varepsilon^{2*}\right) &= f\left[g^{-1}\left(\sigma_\varepsilon^{2*}\right)\right]\frac{dg^{-1}\left(\sigma_\varepsilon^{2*}\right)}{d\sigma_\varepsilon^{2*}} = f\left(S\sigma_\varepsilon^{2*}\right)\frac{dS\sigma_\varepsilon^{2*}}{d\sigma_\varepsilon^{2*}} \\ &= \frac{S}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)\left(S\sigma_\varepsilon^{2*}\right)^{(\underline{\alpha}_\varepsilon+1)}} \exp\left(-\frac{1}{S\underline{\beta}_\varepsilon\sigma_\varepsilon^{2*}}\right) \\ &= \frac{1}{\left(S\underline{\beta}_\varepsilon\right)^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)\sigma_\varepsilon^{2*(\underline{\alpha}_\varepsilon+1)}} \exp\left(-\frac{1}{S\underline{\beta}_\varepsilon\sigma_\varepsilon^{2*}}\right) \end{aligned}$$

Thereby, the prior for σ_ε^{2*} will be changed to $f_{\Gamma\varepsilon^*}^{-1}\left(\underline{\alpha}_\varepsilon^*, \underline{\beta}_\varepsilon^*\right)$ accordingly.

$$p\left(\sigma_\varepsilon^{2*}\right) = \frac{1}{\underline{\beta}_\varepsilon^{*\underline{\alpha}_\varepsilon^*}\Gamma\left(\underline{\alpha}_\varepsilon^*\right)\sigma_\varepsilon^{2*(\underline{\alpha}_\varepsilon^*+1)}} \exp\left(-\frac{1}{\underline{\beta}_\varepsilon^*\sigma_\varepsilon^{2*}}\right)$$

where

$$\underline{\alpha}_\varepsilon^* = \underline{\alpha}_\varepsilon$$

and

$$\underline{\beta}_\varepsilon^* = S\underline{\beta}_\varepsilon$$

Similar for h_ε, with transformed data, the prior of $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ has to be changed to $h_\varepsilon^* \sim f_{\Gamma\varepsilon^*}\left(\underline{\alpha}_\varepsilon^*, \underline{\beta}_\varepsilon^*\right)$. Since $h_\varepsilon^* = g\left(h_\varepsilon\right) = Sh_\varepsilon$, we have $\underline{\alpha}_\varepsilon^* = \underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon^* = S\underline{\beta}_\varepsilon$.

The transformations are linear in this chapter. For nonlinear transformations in the data, refer to Box and Tiao (1992).

Chapter 5

Model Uncertainties and Forecasting: with an Application to Inflation

5.1 Introduction

This chapter focuses on forecasting the UK's inflation rate from a univariate statistical time series modelling perspective. Since the inflation targeting regime was introduced in the 1990s, inflation has become "lower, less volatile and less persistent"(Cogley, 2005) compared with that in the 1970s. In this sense, the inflation rate has been considered to have become easier to forecast. However, Watson and Stock (2007) stress that inflation has become harder to forecast because the forecasts from standard multivariate models, such as the backwards-looking Phillips curve, do not improve over a univariate model. This point was originally made by Atkeson and Ohanian (2001), who found the inflation rate forecasts from the non-accelerating inflation rate of unemployment (NAIRU) models were inferior to the forecast from a naive model[1]. Following Atkeson and Ohanian's argument, Watson and Stock (2007) investigated the postwar U.S. quarterly inflation rate. Strikingly, they found that it is difficult for multivariate forecasts to improve on forecasts made using some univariate models,

[1]In Atkeson and Ohanian (2001), p_t denotes the level of the price index in month t, π_t denotes monthly inflation as measured by $1200\left[\log\left(p_t\right)-\log\left(p_{t-1}\right)\right]$, and π_t^{12} denotes inflation over a 12-month period as measured by $100\left[\log\left(p_t\right)-\log\left(p_{t-12}\right)\right]$. The naive forecast of inflation is given by

$$E_t\left(\pi_{t+12}^{12}-\pi_t^{12}\right)=0.$$

such as the trend-cycle model or integrated moving average model with time-varying parameters. Hence, in the spirit of Watson and Stock (2007), we propose a group of nonlinear forecasting models, the GSTUR model and the SB model, for a univariate U.K. inflation series.

Besides making inference from the proposed nonlinear univariate forecasting models, we also compare the one-step ahead out of sample forecasts from a pool of nonlinear and linear models to that from a benchmark model, the RW model.

After comparing and evaluating the forecasts from univariate statistical forecasting models, we reach the point of forecast uncertainties. As pointed out by Hendry and Ericsson (2001, pp.12), "in practice, forecast uncertainty depends upon what is being forecast, the model used for forecasting, the actual economic process, the current available information and the forecast horizon". According to Clements and Hendry (1998, chapter 7.3, Table 7.1) and Ericsson (in Hendry and Ericsson, 2001, chapter 5), the forecast system is described as an open system, where the model-based forecast errors are normally induced from the following five sources: (1) future changes in the underlying structure of the economy; (2) mis-specification of models; (3) mis-measurement of the data in the base period from which forecasting begins; (4) inaccuracies in the estimates of the model's parameters, and (5) the cumulation of future errors (or "shocks") to the economy.

The 1^{st} source is often recognized as unpredictable future structural breaks or shifts in the underlying economy. This source normally brings a forecast failure to statistical forecasting models (Hendry and Ericsson, 2001, pp.188). In Hendry (in Hendry and Ericsson, 2001, chapter 2, pp.20), the forecast failure is defined as "a significant deterioration in forecast performance relative to the anticipated outcome". In regards to forecasting inflation, "shifts in monetary policy rules alter the fundamentals that drive inflation and therefore also alter its dynamic properties"(Cogley, 2005). In this sense, it is very hard to forecast inflation due to the inherent nonstationarity in the economic system. Questions, such as what kind of economic model is more resilient to breaks, are raised naturally. Analogous to the 1^{st} source of forecast uncertainty, the 2^{nd} and the 3^{rd} sources of forecast uncertainties make the future even more unpredictable. The first three sources of forecast uncertainties are categorized as things that "we do not know what we do not know" (see Hendry and Ericsson, 2001, chapter 2). Whereas, the 4^{th} and the 5^{th} sources of forecast uncertainties are categorized as things that "we do know what we do not know", therefore, are "predictable" and the uncertainties can be quantified. For instance, given a forecasting model, the parameter uncertainties and possible outcomes arising from the shocks, can be presented with

a density forecast. In this regard, by providing point forecasts and density forecasts using the forecasting models, we focus on quantifying forecast uncertainties that arise from the 4^{th} and the 5^{th} sources. To assess the forecasts, or in other words to measure how uncertain the model-based forecasts are, in this chapter we also focus on the point forecast and density forecast evaluations. The point forecasts will be evaluated using the mean squared forecast errors (MSFE) and the density forecasts will be evaluated with the probability integral transforms (PIT).

The 2^{nd} source is identified as uncertainties in model specifications. To reduce the forecast uncertainty that is induced by the 2^{nd} source, a common practice is to combine the forecasts. Combining forecasts has become very popular in the last four decades since the seminal paper by Granger and Bates (1969). Clemen (1989) addresses the point of "forecast accuracy can be substantially improved through the combination of multiple individual forecasts". However, with empirical applications, there is a repeat finding in the literature that simple combination outperforms sophisticated combination methods. This finding is a well known "forecast combination puzzle" (Stock and Watson, 2004). Also, in Clemen (1989), "simple combination methods often work reasonably well relative to more complex combinations". In the Bayesian framework, Bayesian model averaging (BMA) provides "a conceptually elegant means" (Mitchell and Hall, 2007), which accounts for the model uncertainties. Therefore, it is worth evaluating whether a simple averaging is the best combining method compared to BMA. In this chapter, we focus on forecast combinations and see if the BMA approach outperforms the Simple Averaging (SA) approach.

The 3^{rd} source may also bring forecast failure. Watson and Stock (2007) found significant changes in inflation dynamics that "the variance of permanent disturbances to inflation has changed considerably over time, and (the variance) today is at historically low levels". With applications to the U.K. quarterly inflation rate, the introduction of an inflation targeting regime brings an obvious change in monetary policy. Hence, "it is not clear that past experience is a good guide to this [inflation forecast]...and, in turn, [this] probably implies that the error variance... overstate the current uncertainty associated with the inflation rate " (see Poulizac et al. 1996). Taking the National Institute of Economic and Social Research (NIESR)'s inflation forecast for example, NIESR over-estimated the degree of uncertainty associated with its point forecast for the period 1994-2002 (Hall and Mitchell, 2004). Not surprisingly, Mitchell (2005) found a break in the unconditional variance of NIESR's forecast errors around 1993-94. Based on the lesson from NIESR's forecast failure, Hall and Mitchell (2007) propose that historical forecast errors should be monitored regularly by testing struc-

tural breaks at an unknown point, such that a period of history which is informative about the future can be selected for forecasting. In this regard, considering possible mis-measurement of the data in the base period, we carried out the one-step ahead out of sample forecast for 2004Q3-07Q1 using two samples: a period 1957 Q1 to 07Q1 and a later period after the independence of the Bank of England 1999 Q4 to 07Q1. We would like to address the following question, which was also asked in Chatfield (1988): should we simply ignore part of the past data whenever it becomes irrelevant for forecasting purposes?

In summary, the following questions shall be answered in this chapter: Within our selected model space, which statistical model is better in forecasting? Explicitly, are the nonlinear GSTUR model and the Stationary Bilinear (SB) model able to capture the complicated dynamics in the underlying process of inflation, and outperform the simple linear stationary AR(p) models and the Random Walk (RW) model? How do we evaluate the density forecast and point forecast? An immediate question is, how should we combine the forecasts? Compared to a simple averaging, does Bayesian model averaging offer a better solution for combining forecasts? The final question is, when a break occurs in the data sample, which base period should we choose to forecast? Shall we stick to the long historical data set, or alternatively, shall we give up part of the past data set, which is seemingly less relevant to forecasting the future?

Bearing in mind of the five main sources that induce the forecast uncertainties, this chapter is organized as follows: Section 2 reviews the forecast methodologies, which include constructing density forecast, combining forecasts and evaluating point and density forecast. Section 3 then presents the statistical forecasting models. With an application to the UK inflation rates data series, section 4 illustrates the forecast results based on the forecasting models. The forecasts are constructed and evaluated in various ways considering the five main sources inducing the forecast uncertainties. Concluding comments are made in Section 5.

5.2 Forecasting Methodologies

Since forecasting in a Bayesian framework and a Bayesian approach to reducing model uncertainties via BMA have already been reviewed in Chapter 2, in this section, we focus on first the methodologies of constructing and combining forecast in section 5.2.1 and section 5.2.2, then the methods to evaluate point and density forecasts in section 5.2.3. In particular, we focus on reviewing the density forecast and its importance in section 5.2.1

5.2.1 Constructing Forecasts

In the literature, the methodologies of constructing point forecast, interval forecast (Hansen 2006, Chatfield 1993, Christoffersen and Peter 1998) have been well established. Although point and interval forecast meet the basic needs from most professional forecasters, people became more interested in more explicit descriptions of forecast uncertainties. The conventional point forecasts associated with a standard error are no longer satisfactory. Density forecast has gained a lot of attention in recent years (Diebold et al., 1997) because density forecast can serve the above requirement of explicitly quantified forecast uncertainties. For instance, in financial risk management, a density forecast can be utilized when people are more interested in the probability of the realized data exceeding a certain level.

> "A density forecast of the realization of a random variable at some future time is an estimate of the probability distribution of the possible future values of that variable."—Tay and Wallis (2000).

Because the density forecast provides a complete description of the uncertainties of forecast variables, all forecast users can make optimal decisions. Compared with the point forecast, the density forecast has obvious advantages to help forecast users making decisions, when we take the forecast users' loss function into consideration. If the loss function is symmetric, the optimal point forecast is the conditional mean. However, if the loss function is asymmetric, the point forecast based on the conditional mean will be a sub-optimal choice (Christoffersen and Diebold, 1997). However, a density forecast could serve the purpose of providing a probability distribution of all possible outcomes base on the available information. Density forecast has been widely applied in the fields of finance and risk management, as well as in the fields of macroeconomic forecasting. Especially in the last decade, the density forecast has been advanced because of the recent development in Markov Chain Monte Carlo (MCMC), see Gelman (1995) and Diebold et al. (1997). In a Bayesian framework, to obtain density forecasts is straight forward. As long as the posterior simulator is available, model-based density forecasts can be obtained without extra cost via MCMC simulations.

Tay and Wallis (2000) review the importance and usage of density forecast in macroeconomics and finance. We provide a summary here. In macroeconomics, density forecasts are applied to assist central banks to set monetary policy. Taking the inflation density forecast provided by the Bank of England for example, the judgement from the Monetary Policy Committee (MPC) of their views of the balance of risks on

the upside and downside of the inflation forecast are all reflected in the "fan chart"[2]. In quantitative finance, the main focuses are applying density forecast for risk assessments based on forecast users' loss functions, in particular, providing density forecasts of future portfolio values and to tracking certain aspects of the densities (see Diebold et al. 1998). Since it is important to provide accurate and complete probability statements, density forecast, particularly the volatility density forecast and its evaluation have gained remarkable attention. In finance, the most commonly applied forecasting models are ARCH and GARCH models.

To construct a density forecast, numerical simulation methods are applied in most cases. MCMC in the Bayesian framework provides a good solution for simulating forecast distributions. The analytic approach to constructing a density forecast will only be available if standard distributions are used (Tay and Wallis, 2000).

After the forecast densities are constructed, it is very important to present as many features of the forecast density as we can to serve the different needs of all forecast users. Some forecast users might be interested in asymmetries and excess kurtosis, while some forecast users might be interested in some particular high forecast density intervals, or quantiles of the forecast distribution. Therefore, the best way to present a density forecast is by plotting the whole forecast distribution.

To illustrate how density forecasts should be presented, we take the Bank of England's inflation density forecast as an example. With an inflation targeting regime, the objective policy is sometimes expressed as a target range for inflation, whereupon the fan charts are used to report the forecast probability that the future outcome will fall in the target range (Tay and Wallis 2000). In the summary of Britton et al. (1998), the fan charts provided by the Bank of England could be understood in three ways: (1) The final calibration of the distribution represents the Monetary Policy Committee's (MPC) subjective judgements. (2) The central projection of inflation, which is known as the mode of the statistical forecast distribution, represents the most likely outcome based on current knowledge and judgement. The degree of uncertainty can be measured with variance, which is known as the degree of dispersion in the forecast distribution. Also, the degrees of uncertainty are used to reflect how likely it is that future events will differ from the central projection. (3) Third, the degree of skewness in the density forecast distribution exhibits the MPC's view of the balance of risks on the upside and downside of the forecast. When the risk is unbalanced around the central projection, the mean forecast will differ from the mode and the forecast distri-

[2]The plots of the denstiy forecast provided by the Bank of England are also known as the "fan charts".

bution will be asymmetric. If a forecast distribution is positively skewed, the balance of risks will be on the upside, where the mean is higher than the median and the mode. In the Bank of England's report, the degree of skewness is illustrated by the difference between the mean and the mode of a forecast distribution.

In conclusion, to present a density forecast efficiently and make it available for all forecast users, the summary statistics including the mean, mode, median, degree of skewness and variance of a forecast distributions are very important. In this chapter, we plot the one-step ahead forecast distributions to illustrate the model-based forecast densities.

5.2.2 Combining Forecasts

In the forecasting framework, combining forecasts may improve the forecast accuracy. See the seminal paper of Bates and Granger (1969). Also, in Newbold and Granger (1974), they find that regardless of the combination method employed, a small improvement in forecast accuracy results from the addition of a third forecast. The importance of combining forecasts has been briefly reviewed in chapter 2.5. In this section, we provide a detailed review of issues related to forecast combinations, in particular the Bayesian Model Averaging (BMA), with a focus on the density forecasts. We provide, first, a literature review of combining forecasts and, second, two types of methods to combine the forecasts: the Simple Averaging (SA) and the BMA. The algorithms of these two combining methods are provided at the end of this section.

As pointed out by Tay and Wallis (2000), "while a density forecast provides a representation of the uncertainty in a point forecast, its own uncertainty should also be acknowledged and quantified". In a model-based forecasting framework, combining forecasts may provide a solution to reducing the forecast errors that are induced by the model uncertainties. Intensive literature reviews of combining forecast methodologies can be found in Bates and Granger (1969), Clements and Hendry (1998, chap10, pp.228-231), Clemen (1989) and de Menezes et al. (2000). Some non-standard forecast combination methods can be found in Gupta and Wilton (1987) and Li et al. (2001). Clemen (1989) provides a systematic review of combining forecasts and an annotated bibliography of the combining forecast literature. In particular, de Menezes et al. (2000) provide reviews of practical guidelines for the use of combined forecasts and different combining methods. In a focus of density forecast combination, a comprehensive literature review can be found in Hall and Mitchell (2007). For empirical applications, examples can be found in Stock and Watson (2004), where simple combination forecasts, discounted MSFE forecasts, shrinkage forecasts, factor model

forecasts and time varying parameter (TVP) combination forecasts were used. Also, Min and Zellner (1993) applied various Bayesian pooling techniques and non-Bayesian forecast combining techniques to forecast 18 countries growth rates.

In a summary of the literature and using the surveys from de Menezes et al. (2000), there are many well established methods to combine model-based forecasts. The most common idea of combination is giving combination weights to different forecasts, and the forecasts are then combined via a linear weighing vector. The conventional forecast combination methods are listed as follows:

1. Optimal Combining and Optimal Adaptive Combining assuming independence: In these two combining methods, a linear weights vector is constructed by minimizing the variance-covariance matrix of the combined forecast errors. Since the minimization requires the variance-covariance matrix of the combined forecast errors to be properly estimated, the Optimal Combining (OC) approach is not appropriate in practice because the covariance matrix of forecast errors is often non-stationary. The Optimal Adaptive Combining (OAC) approach assumes forecast errors from individual modes are independent and the estimate of the variance-covariance matrix of the combined forecast errors is restricted to be diagonal, comprising just the individual forecast error variances. This OAC method tries to overcome the problem caused by the non-stationary covariance. For details of the variance-covariance approach, refer to Clements and Hendry (1998, pp.229, chap10.2.2).

2. Regression based approach: The regression method is proposed by Granger and Ramanathan (1984), in which the individual forecasts are used as regressors and the weights are obtained via an Ordinary Least Squares (OLS) regression. This approach, is supported by Guerard (1987) and Holmen (1987). However, the approach is not appealed in Clemen (1986), Holden and Peel (1986) and Lobo (1991), who provide empirical evidence favouring the optimal combining approach over the regression based approach. For details of the regression based approach, refer to Clements and Hendry (1998, pp.229, chap10.2.1).

3. Simple Averaging (Clemen and Winkler, 1986): A Simple Averaging (SA) approach to combining forecasts is to simply take the arithmetic average of individual forecasts. The simple averaging has "the virtues of impartiality, robustness and a good track-record in economic and business forecasting" (de Menezes et al., 2000). The SA method has been favoured in the forecast combination literature, because the SA works so well that the SA performs the best or nearly

the best compared with other sophisticated combination methods (see Clemen, 1989 and de Menezes et al., 2000). In some instances, Palm and Zellner (1992) propose to use equal weights instead of using complicated weights to combine forecasts. Some possible explanations of why a SA works well can be found in Holden and Peel (1986). They point out that the weights to combine are potentially unstable, which are often due to the unsystematic changes over time in the variance-covariance matrix of the individual forecast errors. "Under these circumstances, a simple average, although having non-optimal weights, may still give rise to better results than time-varying weights" (de Menezes et al., 2000).

4. Bayesian approaches: Bayesian approaches to combination have been proposed by Morris (1977). Hoeting et al. (1999) provide a detailed introduction of BMA. In a very early paper, Bunn (1985) provides an "outperformance" approach to combine the forecasts, which is a linear combination method with a Bayesian approach. In the outperformance method, the combination weights of each individual forecast can be assessed and revised in a Bayesian manner. The weights are constructed as the probabilities that each individual forecasting model will perform the best on the next occasion, where the probabilities are evaluated based on the models' performances in the past. When data is sparse, the experts judgment will be taken into account to construct the combination weights.

5. Other approaches, such as combining with changing weights can be found in Deutsch et al. (1994). In a recent paper of Li et al. (2001), a quasi-Bayes method (Faria and Souza, 1995) is applied to combine forecast densities.

To sum up, Bunn (1985) found an overall robustness of OAC assuming independence, outperformance and quasi-Bayes probabilities approaches. In Clemen and Winkler (1986), a simple averaging and the Bayesian approach to combination perform well. The success of simple averaging approach has been a mystery in the forecast combination literature. de Menezes (2000) points out, "the simple methods did much better during and immediately after structural changes, however, overall, the advantage of simple averaging approach is not very clear". In this chapter, we focus on comparing a Bayesian approach to combining forecasts based on model probabilities and a simple averaging to combination by taking the arithmetic average of each individual forecasts. The algorithms of BMA and SA are introduced as follows:

Algorithm 1: Forecast Combination: BMA and SA

Since the predictive distributions of each individual forecasting model can be obtained via simulations, these distributions then can be averaged to provide a combined

forecast density. The combined forecast density via BMA $p(y_{t+1} \mid F_t)$ is:

$$p(y_{t+1} \mid F_t) = \sum_{i=1}^{k} p(y_{t+1} \mid F_t, M_i) \Pr(M_i \mid F_t)$$

where $p(y_{t+1} \mid F_t, M_i)$ is the forecast density of model M_i, and $\Pr(M_i \mid F_t)$ is the model probability of M_i calculated as:

$$\Pr(M_i \mid F_t) = \frac{p(F_t \mid M_i)\, p(M_i)}{\sum_{j=1}^{k} p(F_t \mid M_j)\, p(M_j)} \tag{5.1}$$

where $p(M_i)$ is the prior model probability of model M_i. If we are interested in combining k models and presume all the models of interest are equal likely, $p(M_i) = p(M_j)$ for any $j \neq i$, Equation (5.1) will become

$$\Pr(M_i \mid F_t) = \frac{p(F_t \mid M_i)}{\sum_{j=1}^{k} p(F_t \mid M_j)}$$

where $p(F_t \mid M_j)$ is the marginal likelihood of model $M_j : j = 1, \cdots, k$.

Since the forecast distributions of each individual forecasting model is presented as a histogram plot of random samples from the forecast density $p(y_{t+1} \mid F_t, M_i)$, the BMA can be achieved via combining the forecast histograms according to the model probabilities $\Pr(M_j \mid F_t) : j = 1, \cdots, k$.

For instance, in a histogram plot of forecast distribution M_j, if we have 40 bins, denoted as r, within a range of the minimum value 0 and the maximum value 4, the bin density Bin_{j_r}, denoted as $\Pr(Bin_{j_r} \mid F_t, M_j)$ will be the portion of forecast draws that fall within the bin.

If we want to obtain N samples to construct a combined forecast distribution, according to the bin probabilities $\Pr(Bin_{j_r} \mid F_t, M_j) : r = 1, \cdots, 40$ in model M_j with a model probability $\Pr(M_j \mid F_t)$ of M_j, $N_r : r = 1, \cdots, 40$ number of draws should be uniformly sampled from $Bin_{j_r} : r = 1, \cdots, 40$ in the forecast histogram of model M_j:

$$N_r = N \Pr(Bin_{j_r} \mid F_t, M_j) \Pr(M_j \mid F_t)$$

For SA, all the models are treated as being equally likely and equal numbers of random forecast draws are taken from the constituent forecasting models' histograms. The combining algorithms for BMA and SA are as follows:

1. To construct a combined forecast distribution, we would like to obtain N forecast samples in total from the constituent forecasts, by setting the number of bins in all histograms as 40: $r = 1, \cdots, 40$.

2. With a group of constituent forecasting models $M_j : j = 1, \cdots, k$, using the developed posterior simulators, we can get n one-step ahead out of sample forecast samples of y_{t+1} from model M_j according to the forecast density $p\left(y_{t+1} \mid F_t, M_j\right)$. We call these n sample $S_j : y_{j,t+1}^{(1)}, y_{j,t+1}^{(2)}, \cdots, y_{j,t+1}^{(n)}$. If we denote the number of samples that fall in bin Bin_{j_r} as $x_{j,r} : r = 1, \cdots, 40$, the probabilities of Bin_{j_r} in the forecast histogram with respect to model M_j will be $x_{j,r}/n$:

$$\Pr\left(Bin_{j_r} \mid F_t, M_j\right) = \frac{x_{j,r}}{n}$$

3. Combining the forecasts:

   (a) **BMA**: With the obtained marginal likelihood $p\left(F_t \mid M_j\right)$ of $M_j : j = 1, \cdots, k$, model probabilities $\Pr\left(M_j \mid F_t\right) : j = 1, \cdots, k$ are given by the following:

   $$\Pr\left(M_j \mid F_t\right) = \frac{p\left(F_t \mid M_j\right)}{\sum_{i=1}^{k} p\left(F_t \mid M_i\right)}$$

   To construct a combined forecast with a BMA approach, we can uniformly sample $N \Pr\left(M_j \mid F_t\right) \Pr\left(Bin_{j_r} \mid F_t, M_i\right)$ number of forecasts according to each model of interest for $j = 1, \cdots, k$ and each individual model's forecast histograms with bins $r = 1, \cdots, 40$, where N is the total number of forecast samples in the combined forecast distribution.

   (b) **SA**: An SA method is to construct the combined density by considering all models are equally weighted, where $\Pr\left(M_i \mid F_t\right) = \Pr\left(M_j \mid F_t\right)$ for any i and j in the model space. In this instance, we can uniformly take N/k samples from the forecast histogram of $M_j : j = 1, \cdots, k$ to construct the combined forecast distribution.

4. Parameter estimates and other quantities of interest of the combined forecast can be obtained straightforward in the above manner. For example, the BMA estimate of parameter θ, $\widehat{\theta}_{BMA}$ is

$$\widehat{\theta}_{BMA} = \sum_{j=1}^{k} \widehat{\theta}_j \Pr\left(M_j \mid F_t\right)$$

where $\widehat{\theta}_j$ denotes the posterior mean of model M_j. For details, please refer to Hoeting et al. (1999).

If the model space contains many models, model selection using the marginal likelihood is not feasible. A better model selection and forecast combination procedure should be developed. In the literature, using the predictive likelihood for model averaging has been proposed. We can take this into account for future research.

5.2.3 Evaluating Forecasts

This section introduces the methodologies of point and density forecast evaluations, which will be applied to both the forecasts from each individual forecasting model and the combined forecasts.

We begin by introducing some notation:

$\widetilde{y}_{t+i}$ denotes the i step ahead forecast at time t and the available information set is denoted as I_t. Normally, we use the sequence data history up to time t, denoted as F_t, as a representative of I_t. Because $\widetilde{y}_{t+i}$ will be affected by unpredictable future events, or stochastic shocks in the economy system, $\widetilde{y}_{t+i}$ is considered as a random variable following a forecast distribution, denoted as $F_{t,i}(y)$. Conditional on the available information set, the conditional forecast distribution $F_{t,i}(y \mid I_t)$ is:

$$F_{t,i}(y \mid I_t) = prob(\widetilde{y}_{t+i} \leq y \mid I_t)$$

where $prob(\widetilde{y}_{t+i} \leq y \mid I_t)$ denotes the probability of $\widetilde{y}_{t+i}$ being no greater than y given the information available at time t. The corresponding predictive probability density function then is denoted as $f_{t,i}(y \mid I_t)$, where:

$$f_{t,i}(y \mid I_t)\, dy = prob(y \leq \widetilde{y}_{t+i} \leq y + dy \mid I_t)$$

Finally, we denote y_{t+i} as the realized data at the future time $t+i$. In this chapter, we focus on the one-step ahead out of sample forecast, where $i = 1$ in this instance.

How to evaluate the point forecast has been well established in the forecasting literature. For point forecast evaluations, interested readers are redirected to the survey by Wallis (1995), Diebold and Lopez (1996). Surveys of density forecast evaluations can be found in Clements and Smith (2002), Tay and Wallis (2000). Empirical applications of density forecast with its evaluation can be found in both macroeconomics and finance, see Diebold et al. (1998) and Diebold et al.(1999). Other forecast evaluations, such as the evaluation of the interval forecast can be found in Chatfield (1993) and

Christoffersen (1998). Given the review in the literature, we choose standard methods to evaluate both the point and the density forecasts.

Point Forecast Evaluation: MSFE

For more details of point forecast evaluation, see Granger (2001, in Hendry and Ericsson 2001, pp.97-98) and Clements and Hendry (1998, chap3 pp.54-56). We provide a summary here. Formally, the point forecast is obtained by minimizing the forecast users' loss functions $C(\cdot)$. Therefore, the one-step ahead out of sample point forecast are obtained via minimizing the following:

$$\min_{\widetilde{y}_{t+1}} \int_{-\infty}^{\infty} C(\widetilde{y}_{t+1} - y) f_{t,1}(y \mid I_t) dy$$

However, minimizing the above integral directly is difficult. Based on an assumption of symmetric $C(\cdot)$ [3], the one-step ahead point forecast will be just the conditional mean of the predictive probability density function $f_{t,i}(y \mid I_t)$ for linear econometric models, e.g. the stationary AR(p) model and the Random Walk (RW) model. For nonlinear econometric models, such as the Generalized Stochastic Unit Root (GSTUR) model and the Stationary Bilinear (SB) model, the one-step ahead point forecast can be obtained via simulated forecast densities. Details of how to get the one-step ahead point forecast will be addressed with respect to each individual forecasting model in section 5.3.

In this chapter, we apply a conventional method, the Mean Square Forecast Errors (MSFE), to evaluate the point forecast accuracy. Imagine we carried out the one-step ahead forecast for $\widetilde{y}_{t+1}$ when the available information set is I_t, the point forecast under a symmetric loss function is denoted as $\widehat{y}_{t+1}$. At time $t+1$, when the information set is updated to I_{t+1} and the realized y_{t+1} is observed, the forecast error for the first point forecast can be calculated as:

$$e_{t+1} \mid I_t = (\widehat{y}_{t+1} - y_{t+1}) \mid I_t$$

[3] As summarized in Svensson (2002) footnote 7,

> (under the assumption of a linear model of the transmission mechanism) a quadratic loss function corresponds to a mean forecast, an "absolute deviation" loss function corresponds to a median forecast, an "all or nothing" loss function (a so-called Dirac delta function) corresponds to a mode forecast, and a "zone of indifference" loss function corresponds a condition of equality of the probability densities of the forecast at the edges of the zone.

Svensson believes "a majority of an informed and competent MPC would quickly see the advantages of a quadratic loss function and the corresponding mean forecast".

If we generate a number of τ forecasts, the MSFE is calculated as a mean of the sum squared forecast errors:

$$MSFE = \frac{1}{\tau}\sum_{i=1}^{\tau}\left(e_{t+i} \mid I_{t+i-1}\right)^2$$

The smaller the MSFE is, the better the econometric model does in terms of point forecast given a condition that the loss function $C(\cdot)$ is symmetric.

Density Forecast Evaluations: PIT

Different approaches to evaluate the density forecast have been investigated in the literature. In Wallis (2003), a approach of using the chi-squared tests is proposed with an application of the Bank of England's fan charts. West (1996) focuses on the effect of parameter estimation error on the evaluation procedure. West's approach requires a long time series of predictions and realizations. Mitchell and Hall (2004) applied the Kullback-Leibler information criterion (KLIC) to evaluate density forecast with an application to Bank of England fan charts. The combined density is obtained using the KLIC weights. Comparison of these different approaches to evaluate density forecast is a subject of future research.

In this chapter, following the methodologies stated in Clements and Smith(2002), density forecasting can be evaluated using the Probability Integral Transforms (PIT). Diebold et al. (1999) evaluate a bivariate forecasting model with an application of high frequency returns on foreign exchange, by extending the univariate density forecast in Diebold et al. (1998) to a multivariate case. Clements and Smith (2000) extended the density forecast evaluation method in Diebold et al. (1998) and compared among the nonlinear (self-exciting threshold autoregressive models they considered) and linear forecasting models. The PIT approach is not only applicable to model-based density forecasts, but also is operational for subjective (survey-based) density forecasts. For instance, Clements (2006) uses PIT to evaluate the survey of professional forecasters (SPF) probability distribution of inflation and output growth.

The key idea of evaluating the density forecast is that if the density forecast performs well, the cumulated probability up to the realized data should be Identically Independent and Uniformly distributed. This lies in the fact that "if a sequence of realizations corresponds to i.i.d. draws from a fixed density, then the PIT of the realizations with respect to the density are $i.i.d.$ $f_U(0,1)$" (Diebold et al., 1998). Tay and Wallis (2000) addressed "deviations from uniform i.i.d. (in PIT) will indicate that the forecasts have failed to capture some aspect of the underlying data generating process, and serial correlation in the PIT sequence would indicate poorly modelled dynamics,

whereas non-uniformity may indicate improper distributional assumptions, or poorly captured dynamics or both". Diebold et al.(1998) point out that a density forecast that coincides with the data generating process will be optimal in terms of minimizing expected loss whatever users' loss functions. Therefore, this is an advantage of using the PIT approach to testing whether the forecast density closely corresponds to the actual density because we could simply omit the effects of various loss functions. The PIT can also be applied to evaluate multi-step-ahead density forecast, see Diebold et al. (1998).

To evaluate the density forecast, we firstly calculate the PIT from the forecasted samples. Then, a Kolmogorov-Smirnov (KS) test is applied to test if the PIT are $i.i.d. f_U(0,1)$ distributed. The calculation of the PIT can be found in Diebold et al. (1998). Suppose we have a one-step ahead forecast series with τ periods and the forecast densities for y_{t+1} at t is denoted as $p\left(\widetilde{y}_{t+1}\right)$, the PIT z_{t+1} of realized y_{t+1} with respect to $p\left(\widetilde{y}_{t+1}\right)$ is given by:

$$z_{t+1} = \int_{-\infty}^{y_{t+1}} p\left(u\right) du = P\left(y_{t+1}\right)$$

where $P\left(y_{t+1}\right)$ is the forecast probability of $\widetilde{y}_{t+1}$ not exceeding the realized value y_{t+1}. In other words, $p\left(\widetilde{y}_{t+1}\right)$ is the forecast density and z_t is simply the corresponding cumulative density function (c.d.f.) evaluated at the realized value y_{t+1}. Thus, z_t should be within an interval $z_t \in [0,1]$.

According to Diebold et al.(1998), Diebold et al. (1999) and Clements (2006), suppose the unobservable and unknown true forecasting density is denoted as $f\left(\widetilde{y}_{t+1}\right)$, if the forecast density $p\left(\cdot\right)$ equals the true forecast density $f\left(\cdot\right)$, then for $z_t \in [0,1]$, $q_t\left(z_t\right)$ calculated as:

$$\begin{aligned} q_{t+1}\left(z_{t+1}\right) &= f\left(P^{-1}\left(z_{t+1}\right)\right)\left|\frac{\partial P^{-1}\left(z_{t+1}\right)}{\partial z_{_{t+1}}}\right| \\ &= \frac{f\left(P^{-1}\left(z_{_{t+1}}\right)\right)}{p\left(P^{-1}\left(z_{_{t+1}}\right)\right)}. \end{aligned}$$

$q_{t+1}\left(z_{t+1}\right)$ equals to 1. Note that $q_{t+1}\left(z_{t+1}\right)$ is simply the density of $f_U\left(0,1\right)$. Therefore, the evaluation of whether the forecast density $p\left(\cdot\right)$ equals to the true forecast density $f\left(\cdot\right)$ becomes a test of $q_{t+1}\left(z_{t+1}\right)$ being the density of $f_U\left(0,1\right)$. Equivalently, we can just test if the PIT series $Z = \{z_{t+i}, \forall i = 1, \cdots, \tau\}$ from τ recursive forecasts follows an independent identically distributed uniform distribution $(i.i.d. f_U\left(0,1\right))$. If the τ samples of Z are $i.i.d. f_U\left(0,1\right)$, the density forecast $p\left(\cdot\right)$ would be a good representation

of the true underlying forecast density $f(\cdot)$.

Suppose we have a simulated forecast density $p(\widetilde{y}_{t+1})$ with $\widetilde{y}_{t+1}^{(g)} : g = 1, \cdots, S_1$, and the forecast histogram of $\widetilde{y}_{t+1}$ can be plotted using the simulated forecast samples. Suppose the realized value of y_{t+1} at time $t+1$ is 4.5, the c.d.f. evaluated at 4 denoted as $\Pr(\widetilde{y}_{t+1} < 4)$ is 0.4, and the bins are defined with unity intervals, where the bin probability on the interval $[4, 5)$, denoted as $\Pr[\widetilde{y}_{t+1} \in [4, 5)]$, is 0.1 according to the forecast histogram. Then the PIT z_{t+1}, where $z_{t+1} = \Pr(\widetilde{y}_{t+1} < 4.5)$ can be calculated as follows:

$$\begin{aligned} \Pr(\widetilde{y}_{t+1} < 4.5) &= \Pr(\widetilde{y}_{t+1} < 4) + \frac{4.5-4}{5-4} \Pr[\widetilde{y}_{t+1} \in [4, 5)] \\ &= 0.4 + 0.5 \times 0.1 = 0.45 \end{aligned}$$

Various single and joint tests for $f_U(0,1)/f_N(0,1)$ and i.i.d. can be found in Clements and Smith (2000), Clements (2004), Hall and Mitchell (2004), such as KS test, Doornik-Hansen test for $f_U(0,1)/f_N(0,1)$, Ljung-Box tests and Lagrange multiplier (LM) tests for i.i.d., Berkowitz LR test for both $f_U(0,1)/f_N(0,1)$ and i.i.d. In this chapter, we follow the methods in Clements and Smith (2000) and apply the KS test to test $f_U(0,1)/f_N(0,1)$ of the PIT series, Z.

To test if Z is $f_U(0,1)$, we firstly take the inverse c.d.f. transformation of Z to get Z^* ($Z^* \sim F_N^{-1}(Z)$). Then, the test of Z being $f_U(0,1)$ becomes a test of Z^* being $f_N(0,1)$, where the KS test can be applied. The Matlab (2007a) comprises a KS test, in which the null hypothesis for the KS test is that $Z^* \sim f_N(0,1)$ and the alternative hypothesis is that $Z^* \nsim f_N(0,1)$. If the KS test output is 1, we can reject the hypothesis at the 5% significance levels that Z^* has a standard normal distribution. If the KS test output is 0, we cannot reject the null hypothesis that $Z^* \sim f_N(0,1)$. Because the forecasting period τ in this chapter is only 10, the forecast sample size of Z is therefore as small as 10. The inference from the testing results should be made with caution because the KS test power is low subjecting to the small sample size.

To test the i.i.d. in Z, we follow the suggestion in Diebold et al. (1998) via a visualization of the PIT by plotting the correlograms of the Z on the level and its higher powers, which are calculated as $\left(Z - \overline{Z}\right)$, $\left(Z - \overline{Z}\right)^2$, $\left(Z - \overline{Z}\right)^3$ and $\left(Z - \overline{Z}\right)^4$[4].

To sum up, with the methodologies reviewed in section 5.2.1, 5.2.2 and 5.2.3, this chapter, first, constructs density forecasts and point forecasts using the entertained statistical models, then the forecasts are combined using both the SA and the BMA methods. . Finally, the forecast densities, both the density from the constituent

[4] $\overline{Z}$ is the mean of the PIT series Z.

forecasting models and the combined densities, are evaluated with the PIT approach. Correlograms of the PIT on the level and its higher powers are plotted along with the forecast distributions, as well as the realized values.

5.3 Forecasting Models

In this section, the specified statistical forecasting models are presented. Compared to the classical forecasting approach, one distinctive advantage of applying the Bayesian approach to forecasting is that the forecast distributions can be created simultaneously as soon as the estimations were completed. Especially, when the exact forecast density is hard to achieve analytically, with a Bayesian approach, a forecast distribution could be simulated from a large number of random draws. The forecast procedure with an entertained model can be described as follows: First, in each repetition of the sampling, a vector of random draws can be obtained from the joint posterior densities. Then, a forecast can be generated based on the random posterior samples. Finally, when the posterior sampling iterations are completed, the forecast distributions, as well as the posterior distributions, are obtained.

Koop (2003, pp.73) provides details of how to achieve the density forecasts with a use of random draws from the posterior simulators. If θ denotes a vector that consists of all the parameters of interest in model M_j, with the posterior samplers developed in chapter 3 and 4, we can get a random draw $\theta^{(i)}$ from the posterior conditionals $p(\theta \mid M_j, F_t)$, where $F_t = (y_1, \cdots, y_t)'$ denotes a series of historical observations up to t. With MCMC, a one-step ahead forecast of $y_{t+1}^{(i)}$ can be obtained from model M_j and density $p\left(y_{t+1}^{(i)} \mid M_j, F_t, \theta^{(i)}\right)$. When the iteration number i is increased to S, where the chain converge, posterior samples of $\theta : \theta^{(1)}, \cdots, \theta^{(S)}$ and forecast samples of $Y_{t+1} = \left(y_{t+1}^{(1)}, y_{t+1}^{(2)}, \cdots, y_{t+1}^{(S)}\right)'$ can be obtained immediately. Then, the histogram of Y_{t+1} can be plotted as an approximation of the forecast distribution, from where the high forecast density intervals, e.g. 90% density intervals, can be extrapolated. We can choose the mode as the point forecast, which can be evaluated with the MSFE. When the forecast distribution is symmetric, we can also use the mean of the forecast distribution as the point forecast. Because the forecast distribution is constructed from the random posterior draws, the stochastic uncertainty induced by the unobserved errors or shocks, and the parameter uncertainties induced by the parametric models are already accounted for.

Next, we provide the forecasting models' specifications. The model space consists of a number of purely statistical linear models and some non-standard nonlinear models.

The forecast algorithms with respect to each individual model will be introduced. To determine the optimal lag length in the stationary AR(p) model, we performed a pre-selection with a use of the Eviews outputs according to the Akaike information criterion (AIC). Then, in order to compare the stationary AR(p) models with nonlinear models, we estimate the selected stationary AR(p) models with MCMC sampling algorithms in a Bayesian context.

The reason we have chosen the RW model is that Atkeson and Ohanian (2001) argue that the likelihood of accurately predicting a change in inflation using model inflation forecasting models is no better than a coin flip. Also, in Fisher et al. (2002), a question such as "why does inflation behave like a martingale over some periods while at other times it does not" has been asked. Therefore, we include the simplest RW model as a benchmark for model comparison purposes.

GSTUR Forecasting models

$$\nu_{t+1} = y_{t+1} - \delta\,(t+1) - \gamma \tag{5.2}$$

$$\nu_{t+1} = \exp(\alpha_{t+1})\nu_t + \sum_{i=1}^{l} \lambda_i \triangle\, \nu_{t-i+1} + \varepsilon_{t+1} \tag{5.3}$$

$$\alpha_{t+1} = \phi_0 + \sum_{j=1}^{p} \phi_i \alpha_{t-j+1} + \eta_{t+1} \tag{5.4}$$

where $\varepsilon_{t+1} \sim i.i.d.N(0, \sigma_\varepsilon^2)$ and $\eta_{t+1} \sim i.i.d.N\left(0, \sigma_\eta^2\right)$. The deterministic terms are the intercept γ and time trend δ. The density forecast using the GSTUR model can be achieved using: either (a) an algorithm using the posterior draws, or (b) an algorithm using the Auxiliary Particle Filter. Let's denote $\theta = \left(\boldsymbol{\phi}, \sigma_\eta^2, \sigma_\varepsilon^2, \Omega, \boldsymbol{\lambda}\right)'$, then:

Algorithm 1.a: GSTUR Density Forecast Using Posterior Draws

With the posterior simulator developed in chapter 3, according to the GSTUR forecasting model in Equation (5.2, 5.3 and 5.4)

1. $i = 1$, run GSTUR posterior simulator. See chapter 3 for details regarding the sampling of $\theta^{(i)}$. The first S_0 draws are to be discarded.

2. $i = i + 1$when $i > S_0$, we can have the following with $\theta^{(i)}$ according to Equation (5.4):

$$\alpha_{t+1}^{(i)} \mid \underline{\boldsymbol{\alpha}}_{t+1}^{i}, \phi^{(i)} \sim f_N\left(\phi_0^{(i)} + \phi_j^{(i)} \underline{\boldsymbol{\alpha}}_{t+1}^{i}, \sigma_\eta^{2(i)}\right),$$

where $\underline{\boldsymbol{\alpha}}_{t+1}^{i} = \left(\alpha_t^{(i)}, \cdots, \alpha_{t-p+1}^{(i)}\right)'$ denotes the lagged vector of $\alpha_{t+1}^{(i)}$. According

to Equation (5.3):

$$\nu_{t+1}^{(i)} \mid F_t, \alpha_{t+1}^{(i)}, \theta^{(i)} \sim f_N\left(\exp(\alpha_{t+1}^{(i)})\nu_t + \sum_{i=1}^{l}\lambda_i^{(i)} \triangle \nu_{t-i+1}, \sigma_\varepsilon^{2(i)}\right),$$

Then, according to Equation (5.2),

$$y_{t+1}^{(i)} = \nu_{t+1}^{(i)} + \delta^{(i)}(t+1) + \gamma^{(i)}$$

A forecast sample $y_{t+1}^{(i)}$ then is a valid sample from $p\left(y_{t+1}^{(i)} \mid M_{GSTUR}, F_t, \theta^{(i)}\right)$.

3. Do step 2 until $i = S$, where S denotes the total number of draws. Keep the last S_1 draws[5], the conditional mean of the forecast distribution then will be

$$\widehat{y}_{t+1} = \frac{1}{S_1}\sum_{i=1}^{S_1} y_{t+1}^{(i)}$$

The mode and the median of the simulated forecast distribution can also be calculated from the forecast draws.

Algorithm 1.b: GSTUR Density Forecast Using APF

1. First, obtain the estimates of the parameters from the posterior simulator, denoted as $\widehat{\theta}$, and initialize the starting values of $\underline{\boldsymbol{\alpha}}_2$ as a $M \times p$ of zeros matrix to facilitate the Auxiliary Particle Filter $\underline{\boldsymbol{\alpha}}_2^{(g)} = \left(\alpha_1^{(g)}, \cdots, \alpha_{2-p}^{(g)}\right)$, where $g = 1, \cdots, M$

2. $n = 2$ and $F_n = (y_1, y_2)'$, using the APF described in chapter 3, we can have a $M \times 1$ vector of $\alpha_2^{(g)}$ as a random sample from $p\left(\alpha_2^{(g)} | F_2, \underline{\boldsymbol{\alpha}}_2^{(g)}, \widehat{\theta}\right)$. Then, we stack the $\alpha_2^{(g)}$onto its lagged matrix $\underline{\boldsymbol{\alpha}}_2^{(g)}$ to obtain a $M \times p$ matrix $\underline{\boldsymbol{\alpha}}_3^{(g)}$. Hence, with stacking $\alpha_t^{(g)}$ onto $\underline{\boldsymbol{\alpha}}_t^{(g)}$, we can always have an updated lag matrix $\underline{\boldsymbol{\alpha}}_{t+1}^{(g)} = \left(\alpha_t^{(g)}, \cdots, \alpha_{t-p+1}^{(g)}\right)$ to forecast $\alpha_{t+1}^{(g)}$.

3. $n = n + 1$, until $n = t$. Then, we have samples of $\left(\alpha_2^{(g)} | F_2, \underline{\boldsymbol{\alpha}}_2^{(g)}, \widehat{\theta}\right)$, $\cdots$, $\left(\alpha_t^{(g)} | F_t, \underline{\boldsymbol{\alpha}}_t^{(g)}, \widehat{\theta}\right)$.

4. According to Equation (5.2, 5.3 and 5.4),

$$\alpha_{t+1}^{(g)} \mid F_t, \underline{\boldsymbol{\alpha}}_{t+1}^{(g)}, \widehat{\theta} \sim f_N\left(\widehat{\phi}_0 + \widehat{\phi}_j \underline{\boldsymbol{\alpha}}_{t+1}^{(g)}, \widehat{\sigma}_\eta^2\right)$$

[5] $S = S_0 + S_1$

$$\nu_{t+1}^{(g)} \mid F_t, \alpha_{t+1}^{(g)}, \widehat{\theta} \sim f_N\left(\exp(\alpha_{t+1}^{(g)})\nu_t + \sum_{i=1}^{l}\widehat{\lambda}_i \triangle \nu_{t-i+1}, \widehat{\sigma}_\varepsilon^2\right)$$

$$y_{t+1}^{(g)} \mid F_t, \alpha_{t+1}^{(g)}, \widehat{\theta} = \nu_{t+1}^{(g)} + \widehat{\delta}\,(t+1) + \widehat{\gamma}$$

The conditional mean of the forecast distribution $\widehat{y}_{t+1}$ will then be:

$$\widehat{y}_{t+1} = \frac{1}{M}\sum_{g=1}^{M} y_{t+1}^{(g)}$$

where the stochastic uncertainties in the measurement equation σ_ε^2 are already taken into account.

Stationary Bilinear Forecasting model

$$y_{t+1} = (a + b\varepsilon_t)\, y_t + \varepsilon_{t+1}$$

where $\varepsilon_{t+1} \sim i.i.d. f_N\,(0, \sigma_\varepsilon^2)$. Calculate recursively,

$$y_{t+1} = a f_1\,(t-1, b) + f_2(t-1, b) + \varepsilon_{t+1} \tag{5.5}$$

where

$$f_1\,(t-1, b) = \sum_{i=1}^{t}\left[(-b)^{i-1}\prod_{j=1}^{i} y_{t-j+1}\right]$$

$$f_2(t-1, b) = \sum_{i=1}^{t-1}\left[(-1)^{i+1} b^i y_{t-i+1}\prod_{j=1}^{i} y_{t-j+1}\right]$$

In the forecast SB model, a tight prior of $a \sim f_N\,(0.9, 0.1^2)$ is selected.

Algorithm 2: SB model Density Forecast

1. $i = 1$, initialize $a^{(i)}, b^{(i)}, \sigma_\varepsilon^{2(i)}$
2. $i = i + 1$, sample $\left(a^{(i)}, b^{(i)}, \sigma_\varepsilon^{2(i)}\right)$ from $p\,(a, b, \sigma_\varepsilon^2 \mid y)$ until $i > S_0$, since $\varepsilon_t^{(i)}$ can be calculated recursively as the following:

   $$\varepsilon_t^{(i)} = y_t - a^{(i)} f_1\left(t-1, b^{(i)}\right) - f_2\left(t-1, b^{(i)}\right)$$

   according to Equation (5.5), the one-step ahead forecast is then:

   $$y_{t+1}^{(i)} \mid F_t, a^{(i)}, b^{(i)} \sim f_N\left(\left[a^{(i)} + b^{(i)}\varepsilon_t^i\right] \cdot y_t, \sigma_\varepsilon^{2(i)}\right)$$

3. Do step 2 repeatedly until $i = S$. If we denote $S_1 = S - S_0$, the conditional mean of the forecast distribution $\widehat{y}_{t+1}$ will be

$$\widehat{y}_{t+1} = \frac{1}{S_1}\sum_{i=1}^{S_1} y_{t+1}^{(i)}$$

and the $S_1 \times 1$ vector of $y_{t+1}^{(i)} : i = 1, \cdots, S_1$ constructs a simulated forecast distribution.

Stationary AR(p) Forecasting models

The AR(p) model we consider as stationary, where the inverse of characteristic roots all lie outside the unit circle. See Appendix 5.A for details of the calculation issues involved with the stationary AR(p) models, which include the model estimation and marginal likelihood calculation. The forecasting model is:

$$y_{t+1} = \rho_0 + \sum_{i=1}^{p} \rho_i y_{t-i+1} + \varepsilon_{t+1} \tag{5.6}$$

where $\varepsilon_{t+1} \sim i.i.d. f_N\left(0, \sigma_\varepsilon^2\right)$. With different choices of p, a group of stationary AR(p) models are selected. For simplicity, in this thesis, the AR(p) models we investigated are referred to as stationary AR(p) models and denoted as AR(p) hereafter.

Algorithm 3: Forecast with Stationary AR(p) models

1. $i = 1$, give starting values to $\rho^{(i)} = \left(\rho_0^{(i)}, \cdots, \rho_p^{(i)}\right)' 1\left(A\right)$, where $1\left(A\right)$ is the indicator function where the stationary condition is satisfied.

2. $i = i + 1$, sample $\rho^{(i)}$ and $\sigma_\varepsilon^{2(i)}$until $i > S_0$. Then according to Equation (5.6), we have forecast sample $y_{t+1}^{(i)}$ from a normal forecast density:

$$y_{t+1}^{(i)} \mid F_t, \rho^{(i)} \sim f_N\left(\rho_0 + \sum_{j=1}^{p} y_{t-j+1}\rho_j, \sigma_\varepsilon^{2(i)}\right)$$

3. Repeat step 2 until $i = S$. If we denote $S_1 = S - S_0$, the conditional mean of the forecast distribution will be then:

$$\widehat{y}_{t+1} = \frac{1}{S_1}\sum_{i=1}^{S_1} y_{t+1}^{(i)}$$

One issue that emerges here is how to select the lag length p. We can either run the Gibbs sampler for the AR(p) models with different values of p and select those with

the highest model probabilities, or we could estimate a group of AR(p) models with a standard econometric analysis software, e.g. Eviews, and carry out a pre-selecting procedure according to the Akaike information criterion (AIC) and/or the Schwarz information criterion (SIC). This pre-selection then may help us to predetermine with what length of lags that an AR(p) model could possibly receive the highest model probability. Thus, for the purpose of combining density forecasts in later sections, a group of AR(p) models are firstly selected according to AIC using the results from Eviews, then marginal likelihoods of these selected AR(p) models are analyzed using an analytical integration method.

Forecast Random Walk model

$$y_{t+1} = y_t + \varepsilon_{t+1}$$

where $\varepsilon_{t+1} \sim i.i.d. f_N\left(0, \sigma_\varepsilon^2\right)$. The estimate of σ_ε^2 can be obtained by a simple Ordinary Least Square (OLS) method. The one-step ahead point forecast of $\widehat{y}_{t+1}$, conditional mean of the forecast distribution, will just be y_t and the conditional forecast density is:

$$y_{t+1} \mid F_t \sim f_N\left(y_t, \sigma_\varepsilon^2\right)$$

The forecast distribution can be simulated using the above density. Based on the RW forecasting model, the forecast variation is due to the stochastic uncertainties, which come from the variations of the error disturbance σ_ε^2. The marginal likelihood of the RW model can be obtained analytically. See Appendix A for detailed derivation.

5.4 Forecasting the U.K. Quarterly Inflation Rate

The importance of inflation modelling has been introduced in chapter 4.1.1. In the UK, the annual rate of inflation is targeted at 2% based on the Consumer Price Index (CPI). Given the fact that the UK experienced a period of high inflation throughout the 1970's, history looks quite different from today and possibly the future. As Hendry points out (in Hendry and Ericsson 2001, pp.41), "the main problem confronting successful economic forecasting appears to be the inherent nonstationarity of economic data", a question is raised naturally: when we make a forecast, should we forecast from a system with a poor historical record or should we exclude the information that looks "irrelevant" according to our professional judgement and perception. More explicitly speaking, with respect to econometric univariate time series forecasting, to forecast

the UK's inflation rates nowadays, is it appropriate simply ignoring the data before the inflation target regime was introduced? To answer this question, we carried out recursive one-step-ahead out of sample forecasts for 2004Q3-07Q1 using two different samples of data: a period from 1957 Q1 to 2007Q1 with all available data included, and a later period after the independence of the Bank of England, from 1999 Q4 to 07Q1. Because the size of the small sample is only 30, we only provide 10 forecasts given this constraint..

With respect to these two different samples, a group of Generalised Stochastic Unit Root (GSTUR) models, a Stationary Bilinear (SB) model, a Random Walk (RW) model and a group of stationary AR(p) models are firstly fitted to the data and a BDS test is applied to test for dependence in the data series. Then, one-step ahead density forecasts using each individual model are obtained via simulations alongside with plots of the forecast distributions. And then, both the simple averaging approach and the Bayesian Model Averaging approach are applied to combining the forecasts. Finally, point forecasts are evaluated using Mean Squared Forecast Errors (MSFE) and density forecasts are evaluated using the Probability Integral Transforms (PIT).

Someone may aruge that multi-step ahead forecast might be more interesting. In Hendry and Clements (2003), "forecast accuracy should decline as the forecast horizon increases because more innovation errors accrue and predictability falls". For this reason, we only focus on one-step ahead forecast in this thesis.

5.4.1 Data and Methodologies

Fig (5-1) plots the Quarterly UK RPI Inflation series from 1957 Q1 to 2007 Q1. The UK experienced very high inflation in the 1970's and 1980's (see Chapter 4 for reviews). The inflation rates were controlled to vary within a small range after the introduction of inflation targeting regime in 1990s. The first three fourths of samples and the last quarter of samples in the data series behave quite differently in terms of the mean and variance.

Across all entertained forecasting models, a common prior is chosen for parameter $\sigma_{\varepsilon}^{2} \sim f_{\Gamma_{\varepsilon}}^{-1}\left(\underline{\alpha}_{\varepsilon}, \underline{\beta}_{\varepsilon}\right)$, where $\underline{\alpha}_{\varepsilon} = 1$ and $\underline{\beta}_{\varepsilon} = 0.2$. Before using the Bayesian methods for stationary AR (p) models, the Akaike Information Criteria (AIC)[6] embedded in *Eviews 6* is applied to pre-determine the appropriate length of the distributed lag. The inflation is regressed on its own lags up to 8 quarters, and an AR(7) model is selected with the smallest AIC for the period 1957Q1-2007Q1. Likewise, an AR (5) model

[6]See Judge et.al. (1985, pp. 242-247) of other Criteria methods for model selections.

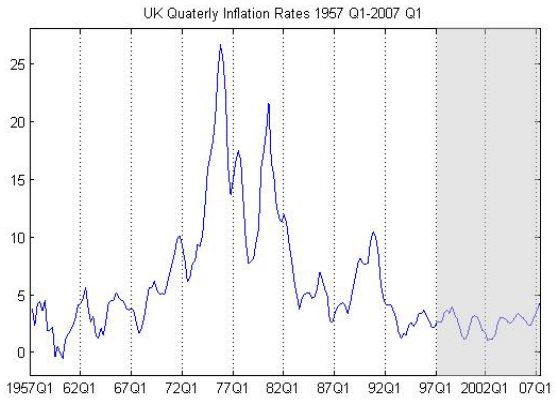


Figure 5-1: Plot of UK inflation series: 1957 Q1- 2007 Q1

is selected for the sample period 1999Q4-2007Q1. Since AR(4) model is commonly applied for inflation modelling and examples can be found in Watson and Stock (2007) and Atkeson and Ohanian (2001), we also include a stationary AR(4) model as a model of interest for empirical applications using both the full range of and a sub-sample of data, where the small sample consists of the last 30 observations.

5.4.2 Application with Full Sample: 1957 Q1-2007Q1

BDS test

To motive our applications of the proposed nonlinear models, we executed a BDS test to investigate how well the model fits the data before constructing a one-step ahead out of sample forecast.

The BDS test is designed to test for time based dependence in a series (Broock et al. 1996). The testing hypothesis are constructed as follows:

H_0 : a time series sample comes from a data generating process that is identically independent distributed (i.i.d.)

H_1 : not specified (a variety of alternatives, e.g. linear dependence, non-linear dependence, or chaos).

In practice, the BDS test is applied to test for residuals i.i.d. In other words, by fitting a model (M_1) to the data observations and obtain a residual series, we remove a deterministic structure M_1 from the system. The failure of the residual series to pass the BDS test may indicate some unspecified serial dependence, which cannot be removed by the deterministic structure M_1.

Table (5.1) provides the BDS test results using the full sample of data with 201

Table 5.1: BDS Ttest Using the Full Sample

	m=2		m=3		m=4		m=5	
	BDS-stat	P-value	BDS-stat	P-value	BDS-stat	P-value	BDS-stat	P-value
GSC0	.046144	0.000	.078472	0.000	.103235	0.000	.123907	0.000
GSCT0	.045907	0.000	.078682	0.000	.103801	0.000	.124351	0.000
GS0	.043522	0.000	.074289	0.000	.096235	0.000	.115533	0.000
GST0	.044224	0.000	.074919	0.000	.097112	0.000	.116718	0.000
GSC1	.027257	.0001	.054054	0.000	.078525	0.000	.103392	0.000
GSCT1	.026876	.0001	.052850	0.000	.078087	0.000	.103848	0.000
GS1	.025503	.0002	.052146	0.000	.074528	0.000	.097461	0.000
GST1	.025484	.0001	.051559	0.000	.074781	0.000	.098408	0.000
GSC4	.038792	0.000	.072671	0.000	.099772	0.000	.120475	0.000
GSCT4	.039425	0.000	.074025	0.000	.102814	0.000	.124252	0.000
GS4	.033664	0.000	.068579	0.000	.096372	0.000	.115566	0.000
GS4T	.033889	0.000	.065289	0.000	.089961	0.000	.108835	0.000
GSC5	.039593	0.000	.073206	0.000	.098648	0.000	.115006	0.000
GSCT5	.039766	0.000	.073679	0.000	.099133	0.000	.115336	0.000
GS5	.032502	0.000	.068268	0.000	.094887	0.000	.110804	0.000
GST5	.032641	0.000	.064483	0.000	.087494	0.000	.103101	0.000
GSC7	.027008	0.000	.062663	0.000	.086686	0.000	.100439	0.000
GSCT7	.031947	0.000	.069055	0.000	.094315	0.000	.108160	0.000
GS7	.032393	0.000	.069564	0.000	.095733	0.000	.110186	0.000
GST7	.027116	0.000	.065359	0.000	.091859	0.000	.106564	0.000
SB	.076227	0.000	.131664	0.000	.168129	0.000	.193713	0.000
RW	.054431	0.000	.093802	0.000	.120774	0.000	.142111	0.000
AR(7)	.040234	0.000	.083968	0.000	.116450	0.000	.134158	0.000
AR(4)	.037636	0.000	.072989	0.000	.104475	0.000	.129539	0.000

observations. The dimension m in the BDS test varies from 2 to 5. The models selected are: 1. GSTUR class models with different specifications in drift γ, deterministic time trend δ, and different lag length l, 2. Stationary Bilinear model, 3. Random Walk model, 4. stationary AR (7) and stationary AR(4) model.

From Table (5.1), after various deterministic structures are removed, the BDS test strongly rejects the hypothesis that the residuals are independent. This indicates the underlying process of inflation has complicated dynamics, which could not be fully captured by any of the above proposed models. Therefore, questions of nonlinear dynamics in inflation series, such as structural breaks, time varying cycles, are open for further research.

Results from Constituent forecasting models

In this section, we construct density forecast for 24 models, which are GSTUR with constant, GSTUR with trend, GSTUR with constant and trend, and GSTUR without constant or trend at lag length of 0,1,4,5,7 respectively, SB model, RW model, and stationary AR(p) model with p selected as 4 and 7, respectively. The density forecasts are represented graphically as a set of prediction intervals covering 10, 20,..., 90 percent of the probability distribution, of lighter shades for the outer bands. Equivalently the boundaries of the bands are the 5th, 10th,...,95th percentiles. Each pair of band covers 10% of the forecast distribution. If the risk is unbalanced, the same color bands are not of equal width, which represents unequal probability intervals. The central projection is the deepest blue since it associates with the mode of the forecast distribution. For unbalanced risks, the mean and median may not be in the deepest blue band. With multiple-step ahead forecast, the forecast intervals "fan out" as the forecast horizon increases. In this chapter, since we focus the one-step ahead forecast only, the fan out effects are not so significant.

We present the descriptive statistics of the forecast distributions in tables, which include the mean, mode, median, variance, 95% percentile of the forecast distributions. Also, conditional on available information at time t, I_t, the marginal likelihoods of the specified models are presented. With lag length fixed at 7, forecast results of GSTUR with constant, GSTUR with trend, GSTUR with constant and trend, GSTUR without constant or trend are provided in Table (5.2, 5.3,5.4 and 5.5). Also in the text, we present the forecast results of stationary AR(7), stationary AR(4), RW and SB model in Table (5.6, 5.7, 5.8 and 5.9). Forecast results from other 16 differently specified GSTUR class of models are presented in Appendix 5.B.

Table (5.10) presents the MSFE for all entertained models. Selected with lag length at 7, comparing amongst the GSTUR class of models, a model with a constant receives the highest marginal likelihood and provides the best point forecast with the smallest MSFE according to Table (5.10).

The evaluation of density forecast, using a KS test testing whether the PITs from the forecast density follows U(0,1)/N(0,1), is presented in Table (5.11). The values of 0 indicate that the null hypothesis of Z, the PITs are U(0,1) distributed, cannot be rejected by the KS test, while the values of 1 indicate that the uniformity in Z is rejected. According to Table (5.11), PITs from the forecast GSTUR models with a trend, forecast GSTUR models with constant and trend could not pass the KS test. The uniformity of the PITs are rejected in GSTUR models with T and CT. It worth mentioning that KS test has low power with small sample sizes. However, in this case,

Table 5.2: Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=7, with Cons

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4473	3.4641	3.4678	2.3629	0.3503	6.4170	-352.3575
05Q1	3.17	3.6497	3.4105	3.6404	2.3106	0.6555	6.5878	-353.0102
05Q2	3.01	3.0566	2.9506	3.0788	2.3005	0.0463	6.0156	-354.2753
05Q3	2.77	2.8924	3.1352	2.9094	2.2411	-0.1031	5.8153	-354.7319
05Q4	2.38	2.7027	2.6055	2.7053	2.2093	-0.2603	5.6205	-355.5625
06Q1	2.39	2.4528	2.1412	2.4615	2.3062	-0.5945	5.3539	-355.3517
06Q2	2.93	2.5122	2.4391	2.5308	2.2691	-0.4665	5.4442	-356.4967
06Q3	3.44	3.2976	3.4792	3.3287	2.2441	0.3076	6.1903	-357.4588
06Q4	3.99	3.8949	3.8879	3.8882	2.2007	1.0022	6.8075	-358.1957
07Q1	4.55	4.2758	4.1319	4.2646	2.1512	1.4102	7.1587	-359.5171

Table 5.3: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with Trend

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4579	3.7104	3.4859	5.0357	-0.9718	7.8211	-353.1912
05Q1	3.17	3.6604	4.3045	3.6745	4.8655	-0.7848	7.9603	-353.3196
05Q2	3.01	3.0333	3.0516	3.0747	4.8347	-1.3921	7.3122	-354.7268
05Q3	2.77	2.9140	2.3840	2.9129	4.9697	-1.4481	7.2283	-355.2979
05Q4	2.38	2.6934	2.1711	2.7160	4.8236	-1.7071	6.9173	-355.7200
06Q1	2.39	2.4618	3.2145	2.5098	4.7927	-1.9426	6.7368	-357.0372
06Q2	2.93	2.4877	2.2645	2.5214	4.8843	-1.9635	6.7197	-356.9981
06Q3	3.44	3.3093	3.3314	3.3162	4.7115	-0.8873	7.4957	-358.2469
06Q4	3.99	3.9386	4.2814	3.9692	4.6509	-0.3474	8.1169	-359.1952
07Q1	4.55	4.2630	4.1427	4.2527	4.6849	-0.0208	8.5130	-359.5879

Table 5.4: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with Cons and Trend

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3710	3.0160	3.3726	5.1045	-1.0801	7.7927	-371.5125
05Q1	3.17	3.5712	3.3444	3.5603	4.8717	-0.7559	7.9076	-371.6482
05Q2	3.01	2.9770	2.9319	2.9936	5.0968	-1.5731	7.3929	-373.0897
05Q3	2.77	2.8352	2.6938	2.8375	4.8633	-1.5448	7.2819	-373.7779
05Q4	2.38	2.6127	2.7139	2.6177	4.8633	-1.7980	6.9545	-374.3405
06Q1	2.39	2.3813	2.6158	2.3939	4.8722	-1.9687	6.8146	-375.0120
06Q2	2.93	2.4049	2.5049	2.4214	4.6854	-1.8906	6.6591	-376.3249
06Q3	3.44	3.2422	3.1517	3.2371	4.6031	-0.9151	7.4469	-376.6760
06Q4	3.99	3.8293	3.2524	3.8079	4.7046	-0.4466	8.1211	-377.4532
07Q1	4.55	4.1903	3.7276	4.1493	4.8626	-0.1740	8.5739	-377.6779

Table 5.5: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with NO CT

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3019	3.2785	3.2873	1.3394	1.0434	5.6003	-371.6062
05Q1	3.17	3.5136	3.4988	3.5027	1.3256	1.2888	5.7846	-372.8042
05Q2	3.01	2.9127	2.9394	2.9143	1.3573	0.6187	5.1845	-373.9598
05Q3	2.77	2.7799	2.8126	2.7943	1.3070	0.5191	5.0124	-374.0565
05Q4	2.38	2.5471	2.7391	2.5524	1.3011	0.2777	4.7503	-374.9762
06Q1	2.39	2.3026	2.3909	2.2996	1.2759	0.0885	4.5290	-376.0596
06Q2	2.93	2.3468	2.3034	2.3451	1.2611	0.1315	4.5934	-376.5732
06Q3	3.44	3.2019	2.9417	3.2035	1.2965	0.9976	5.4668	-377.4902
06Q4	3.99	3.8128	3.7586	3.8165	1.3260	1.5399	6.0955	-378.3621
07Q1	4.55	4.1529	4.2001	4.1578	1.4009	1.8331	6.4591	-379.0650

Table 5.6: Full Sample:Forecast of Inflation Rates with AR(7)

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	MLint
04Q4	3.41	3.5188	3.4519	3.5203	1.2499	1.2858	5.7163	-286.9884
05Q1	3.17	3.6757	3.2703	3.6714	1.2104	1.5426	5.8484	-287.9693
05Q2	3.01	3.1282	3.3185	3.1270	1.2155	0.9724	5.2735	-289.0666
05Q3	2.77	2.9431	3.1763	2.9455	1.2574	0.7450	5.1688	-290.0434
05Q4	2.38	2.8081	2.5642	2.8035	1.2384	0.6207	4.9820	-291.0267
06Q1	2.39	2.5114	2.4669	2.5139	1.2001	0.2992	4.6453	-292.0586
06Q2	2.93	2.5605	2.3672	2.5590	1.1947	0.4282	4.7107	-293.0298
06Q3	3.44	3.3807	3.0365	3.3769	1.2002	1.2277	5.5048	-294.0569
06Q4	3.99	3.9430	3.7040	3.9436	1.1763	1.7919	6.0549	-295.0184
07Q1	4.55	4.3197	4.3623	4.3236	1.1661	2.1782	6.4364	-295.9765

Table 5.7: Full Sample:Forecast of Inflation Rates with AR(4)

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	MLint
04Q4	3.41	3.4166	3.4363	3.4025	1.4371	1.1142	5.8125	-305.3242
05Q1	3.17	3.6858	3.5281	3.6762	1.4582	1.3080	6.0254	-306.3953
05Q2	3.01	3.1957	3.3044	3.1850	1.3854	0.9364	5.5151	-307.5587
05Q3	2.77	3.0309	2.9023	3.0136	1.3824	0.7251	5.2970	-308.6347
05Q4	2.38	2.7902	2.7060	2.7832	1.4129	0.4920	5.1520	-309.7232
06Q1	2.39	2.3623	2.2889	2.3628	1.4466	-0.0301	4.7678	-310.8564
06Q2	2.93	2.5427	2.4940	2.5376	1.4036	0.2687	4.8632	-311.9160
06Q3	3.44	3.3500	3.2491	3.3509	1.3609	1.0885	5.6476	-313.0278
06Q4	3.99	3.8104	3.7363	3.7985	1.4301	1.4856	6.1941	-314.0855
07Q1	4.55	4.3287	4.5431	4.3424	1.3795	2.0294	6.5985	-315.1494

Table 5.8: Full Sample:Forecast of Inflation Rates with RW

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	MLint
04Q4	3.41	3.0622	3.1431	3.0621	1.7639	0.4768	5.6970	-326.3807
05Q1	3.17	3.3812	3.2015	3.3809	1.7659	0.7460	5.9757	-327.6197
05Q2	3.01	3.1427	3.0835	3.1522	1.7383	0.5364	5.7155	-328.8455
05Q3	2.77	2.9569	3.3010	2.9656	1.7503	0.3468	5.5289	-330.0599
05Q4	2.38	2.7248	2.9000	2.7137	1.7352	0.1650	5.2896	-331.2808
06Q1	2.39	2.3576	2.4409	2.3599	1.6912	-0.1798	4.9368	-332.5264
06Q2	2.93	2.3538	2.1036	2.3551	1.7082	-0.1845	4.9080	-333.7261
06Q3	3.44	2.9042	3.0401	2.8988	1.7034	0.3594	5.4745	-335.0075
06Q4	3.99	3.4046	3.4777	3.4000	1.7146	0.8560	5.9740	-336.2780
07Q1	4.55	3.9267	3.7495	3.9250	1.6873	1.4048	6.4646	-337.5590

Table 5.9: Full Sample:Forecast of Inflation Rates with SB

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.0554	3.2458	3.0821	1.4361	0.6854	5.3396	-247.5789
05Q1	3.17	3.3573	3.5173	3.3725	1.4668	0.9813	5.7190	-249.5266
05Q2	3.01	3.0634	3.0549	3.0685	1.4379	0.6958	5.3498	-249.5073
05Q3	2.77	2.9570	2.8864	2.9680	1.4495	0.6104	5.2664	-250.8939
05Q4	2.38	2.6919	2.2472	2.6782	1.4172	0.3350	5.0894	-251.8722
06Q1	2.39	2.2746	2.0813	2.2584	1.4262	-0.0725	4.6025	-253.4594
06Q2	2.93	2.3653	2.5430	2.3493	1.4176	0.0626	4.7416	-253.0735
06Q3	3.44	2.9243	2.9951	2.9467	1.3828	0.6101	5.2073	-254.0916
06Q4	3.99	3.4462	3.2617	3.4476	1.3347	1.1538	5.6786	-255.2687
07Q1	4.55	3.9843	3.8889	3.9912	1.4148	1.6927	6.3588	-256.5976

Table 5.10: Full Sample: MSFE of Statistical Forecasting Models

GSC:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	SB
0.1331	0.0706	0.058	0.0618	0.0578	0.1528
GST:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	RW
0.1302	0.0746	0.0553	0.0609	0.0605	0.1535
GSCT:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	AR(4)
0.1386	0.0792	0.0636	0.0627	0.0629	0.0779
GSnoCT:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	AR(7)
0.1539	0.0863	0.0683	0.073	0.0692	0.0694

we should not worry because still reject the null given the small sample size.

Table 5.11: Full Sample: KS Test for PIT i.i.d.U

GSC:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	SB
0	0	0	0	0	0
GST:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	RW
1	1	1	1	1	0
GSCT:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	AR(4)
1	1	1	1	1	0
GSnoCT:$l=0$	$l=1$	$l=4$	$l=5$	$l=7$	AR(7)
0	0	0	0	0	0

In the text, the density forecast plots of GSTUR with constant and lag length of 7, GSTUR with trend and lag length of 7, stationary AR(7), RW, and SB model are presented. Other results are shown in Appendix 5.B.

According to Clements (2004), we can simply plot the correlograms in the sample of $\{z_t - \overline{z}\}$, $\overline{z} = n^{-1}\sum_{t=1}^{n} z_t$ and of powers of this series $\left\{(z_t - \overline{z})^i\right\} : i = 2, 3, 4$ to test independence (*i.i.d.*) in Z on both levels and in higher moments. Correlograms of PITs and its higher powers, together with a scatter plot of the PITs are presented on the side of the density forecast plots. Forecast density plots of other 19 models can be found in Appendix 5.B. All the forecast densities are evaluated via the PIT approach with a KS test, correlogram plots and scatter plots.

According to Figures 5.d(2) and 5.e(2), the PITs from RW and SB exhibit strong serial correlations, while the PITs from the selected GSTUR model with a constant, GSTUR model with a trend and the stationary AR(7) model exhibit smaller serial correlations on the levels and higher moments. Thus, serial correlation in the PIT sequence indicates that the inflation dynamics may not be well represented by the RW and SB models.

Once again, because the forecast sample is small, the independency of z_t using the correlograms should be inferenced with caution.

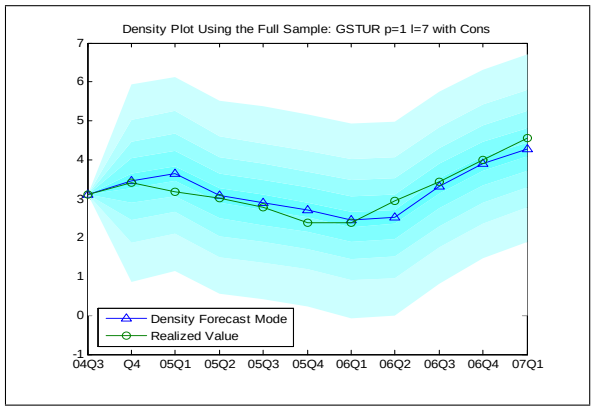


Fig 5.a(1) GSTUR p=1 l=7 with Cons

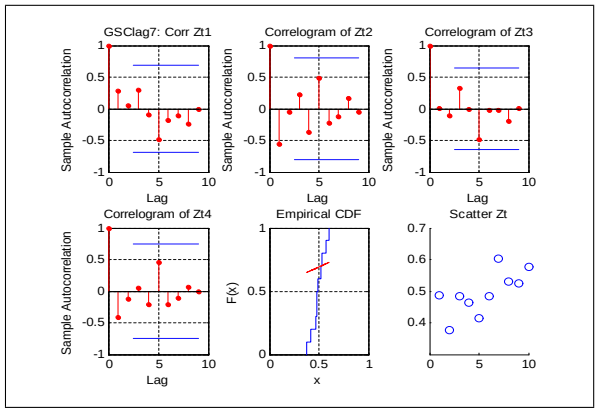


Fig 5.a(1) Correlogram of PITs and higher powers GSTUR p=1 l=7 with Cons

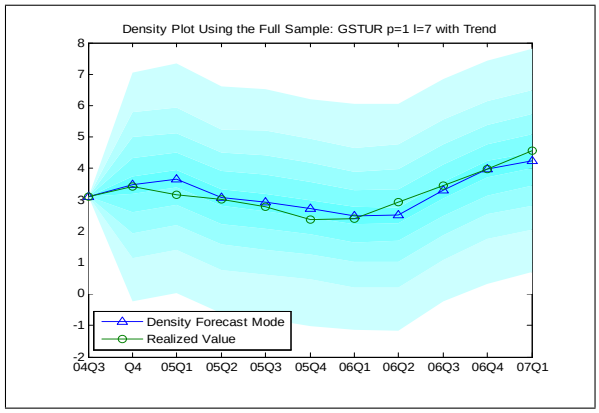


Fig 5.b(1) GSTUR p=1 l=7 with Trend

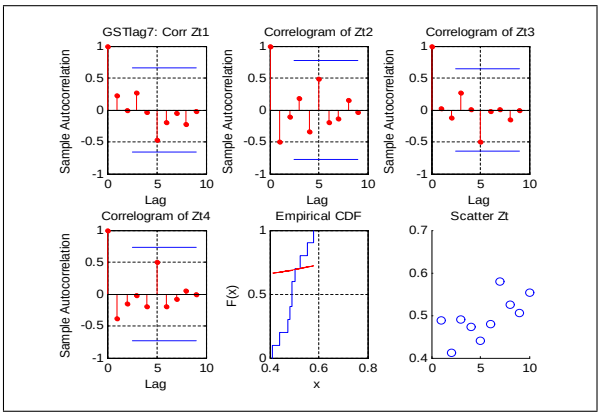


Fig 5.b(2) Correlogram of PITs and higher powers GSTUR p=1 l=7 with T

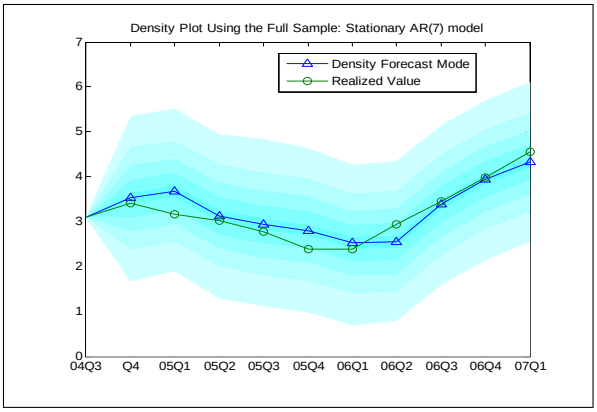


Fig 5.c(1) Stationary AR(7)

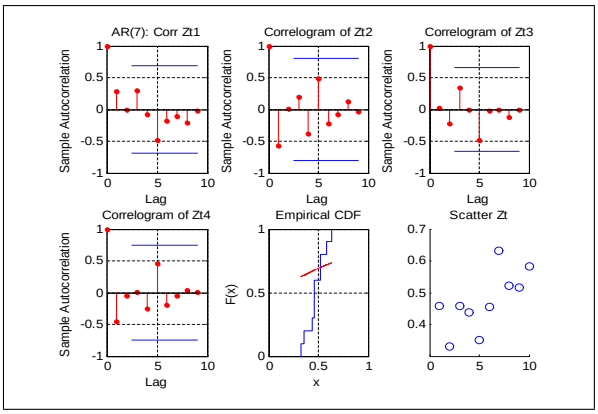


Fig 5.c(2) Correlogram of PITs and higher powers Stationary AR(7)

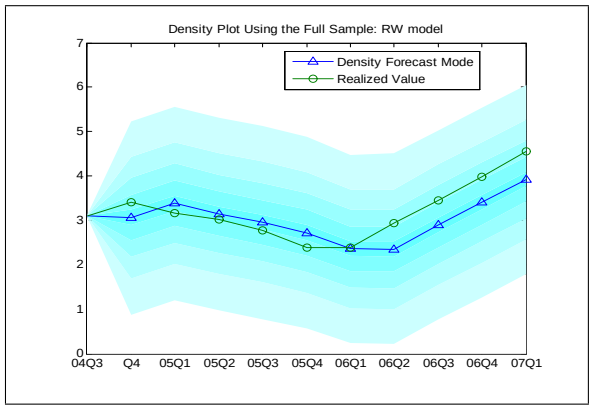


Fig 5.d(1) RW model

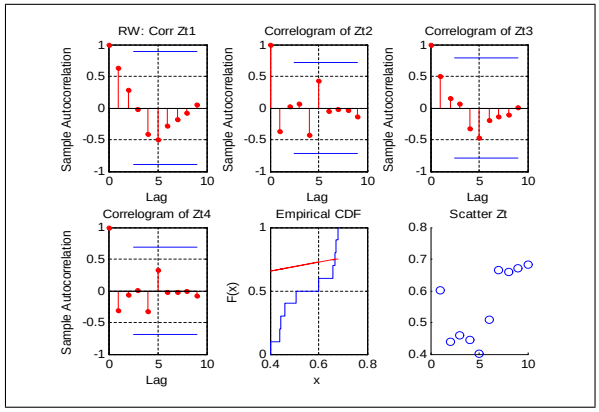


Fig 5.d(2) Correlogram of PITs and higher powers RW

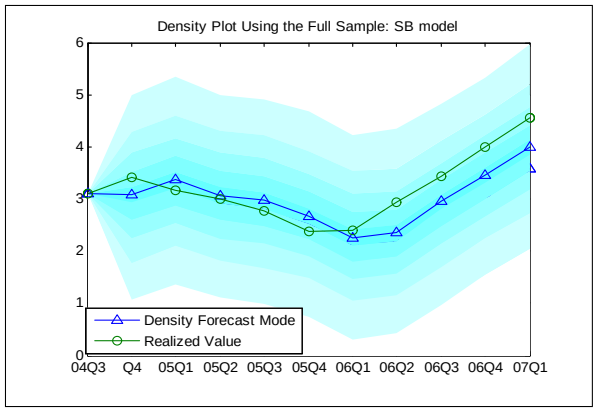


Fig 5.e(1) SB model

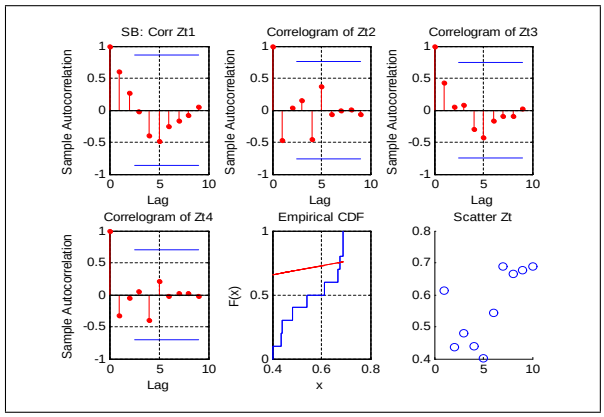


Fig 5.e(2) Correlogram of PITs and higher powers SB model

From Figure 5.a(1) and Figure 5.b(1), a GSTUR model with trend generates larger forecast variances than a model with a constant. From the forecast results using the GSTUR with a constant and trend, the point forecast accuracy is not improved over the simpler models, such as the GSTUR without constant or trend. Moreover, the forecast variances from the GSTUR models with constant and trend are 3 times larger than those from GSTUR models with only a constant. The large forecast variances may be due to the over parameterization in the GSTUR model when both the constant and trend parameters are included. With the same lag length, GSTUR models with a constant and a trend receive lower marginal likelihoods compared with other specifications in the constant and trend. Therefore, model probabilities of models specified with a deterministic trend and a constant are very low. Therefore, it is not appropriate to include models specified with a constant and trend in the forecast averaging.

Across the 10 forecast periods, although the SB model receives the highest marginal likelihood among all 24 entertained models, the point forecast results from the SB model is not satisfactory based on MSFE value. With respect to the point forecast plots, the performances of the SB and the RW forecasting models are similar see Figure 5.d(1) and 5.e(1). This forecast similarity between the SB and RW forecasting models may be due to the tight prior of a we elicited in the SB, which compels the SB dynamics to exhibit like a RW process. Under the elicitation of tight priors, the SB model is not as resilient as the GSTUR model.

The stationary AR(7) receives the highest marginal likelihood compared with a stationary AR(4) and RW model. Also, the AR(7) provides the smallest value of MSFE. Comparing between the AR(7) and GSTUR models with lag length at 7, the

Table 5.12: Full Sample: Simple Averaging Density Forecasts

		Density forecast				95% percentile	
	TrueV	mean	mode	median	var	2.5%	97.5%
04Q4	3.41	3.5188	3.4519	3.5203	1.2499	1.2858	5.7163
05Q1	3.17	3.6757	3.2703	3.6714	1.2104	1.5426	5.8484
05Q2	3.01	3.1282	3.3185	3.1270	1.2155	0.9724	5.2735
05Q3	2.77	2.9431	3.1763	2.9455	1.2574	0.7450	5.1688
05Q4	2.38	2.8081	2.5642	2.8035	1.2384	0.6207	4.9820
06Q1	2.39	2.5114	2.4669	2.5139	1.2001	0.2992	4.6453
06Q2	2.93	2.5605	2.3672	2.5590	1.1947	0.4282	4.7107
06Q3	3.44	3.3807	3.0365	3.3769	1.2002	1.2277	5.5048
06Q4	3.99	3.9430	3.7040	3.9436	1.1763	1.7919	6.0549
07Q1	4.55	4.3197	4.3623	4.3236	1.1661	2.1782	6.4364

GSTUR class of models improved over the AR(7) model in terms of better accuracy of point forecasts (refer to Table 5.10).

Combining Forecasts

With BMA methodology to combine forecasts, the SB model dominates the combined forecast density because the SB model obtains the highest marginal likelihood throughout the 10 forecast periods. However, the combined density does not improve upon individual forecasting models due to the poor forecasting performance of the SB model.

We then focus on the simple averaging method to combine forecasts. To do a simple averaging, we firstly selected the forecast densities with PITs passed both the KS test and the i.i.d. evaluation, and also providing the smallest MSFE. The six models we have chosen to combine are GSTUR model with constant only, lag length with 4 and 7, GSTUR model without constant or trend, lag length with 4 and 7, the stationary AR(4) and stationary AR(7).

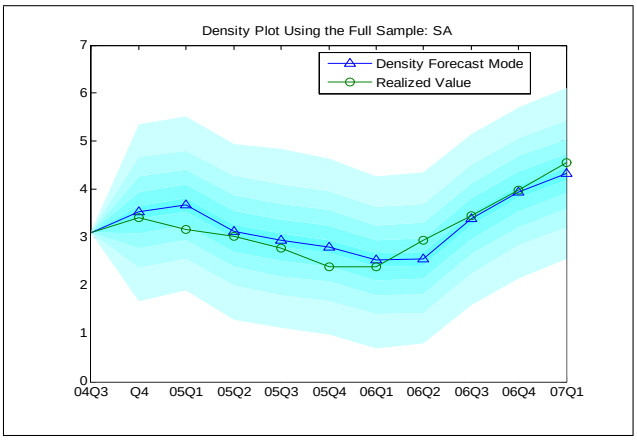


Fig 5.f(1) SA of Forecast Density Plot

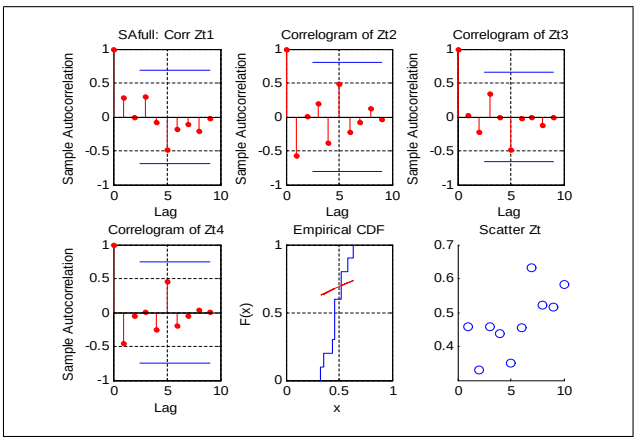


Fig 5.f(2) Corr Plots of PIT from Averaged Forecast Density

Figure 5.f(1) plots the combined forecast density with a simple averaging method and Figure 5.f(2) plots the correlograms of the PIT series and higher moments from the combined density. According to the KS test results, the U(0,1) of PIT is not rejected and there is no strong serial correlation in the PITs. The MSFE of combined points forecast using the combined density is 0.705, which slightly outperformed the stationary AR(4) model. In this case, combining the proposed 6 models could not provide a better point forecast. This may be due to the fact that the constituent forecasting models follow similar structures and could not be disentangled from the other models. The identification problems may result in low point forecast accuracies when we combine the forecasting models.

5.4.3 Application with Small Sample: 1999 Q4-07Q1

Figure (5-2) plots the estimated stochastic roots by fitting the inflation data series with a GSTUR model without a constant or trend, with a lag length selected at 7. The $*$ on Figure (5-2) shows the roots jump above one corresponding to the inflation rate data series. The pattern of the stochastic roots before 1993 looks very different from that after 1993. Before 1993, the stochastic roots jump above one frequently. However, the roots stayed below unity after 1993. The result from estimating a GSTUR model with an application of the UK inflation provides good evidence of a changing persistence in the inflation underlying process, which is consistent with Watson and Stock (2007): "the variance of permanent disturbances to inflation has changed considerably over time". Also, on the plot of the estimated roots, a clear break in 1991 can be observed. Therefore, we are motivated to apply the GSTUR class models, the SB model, RW and the stationary AR(p) models to forecast a period from 2004 Q4 to 2007Q1 using a small sample after the independence of Bank of England, in addressing the following questions: Does linear AR(p) models produce better point forecast than the complicated nonlinear GSTUR models when the data looks linear? Instead of using the full sample, if we apply the same statistical forecasting models to a small sample to forecasting the same period, from 2004 Q4 to 2007Q1, does a linear AR(p) model outperform other specified forecasting models? The small sample data period is from 1999Q4 to 2007Q1 with a sample size of 30. In this time period, inflation targeting policy was implemented and the inflation rate stayed below 6%.

The first out-of-sample forecasting sample, the forecast of the inflation rates in 2004 Q4, is obtained based on a period from 1999 Q4 to 2004 Q3, with a sample size of 20. The one-step ahead forecasting is carried out repeatedly 10 times with a rolling window to achieve 10 forecasts for the period 2004 Q4 to 2007Q1.

Before using the Bayesian methods for stationary AR (p) models, we used Eviews 6.0 for AR(p) model's estimation. Then, a stationary AR(5) is selected according to the smallest AIC. We also include a stationary AR(4) model as a model of interest. In summary, the entertained models are GSTUR class models without a constant or trend at lag length of 0,1,4, and 5, SB model, RW model, stationary AR(5) and AR(4) model. In the SB model, a tight prior of the parameter $a \sim f_N\left(1, 0.1^2\right)$ is elicited.

BDS test

With the small samples, we adopt the bootstrapped p-values. Bootstrapping the BDS test gives better size properties but does not affect the size-adjusted power of a test.

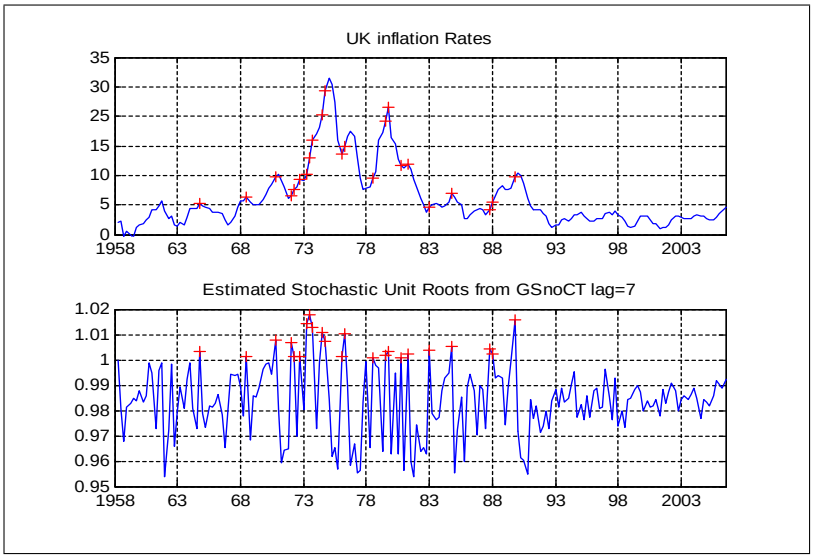


Figure 5-2: Simulated Roots Using the GSTUR model with No Constant or Trend with lag length at 7

The number of repetitions in the Bootstrapping is chosen as $1,000$.

From Table (5.13), removing both the GSTUR and AR(p)[7] structures from the series may provide *i.i.d.* residuals according to the bootstrapped p-values. However, removing the SB and RW structures, dependences still exist in the residuals, where the null hypothesis of *i.i.d.* in the residual series is rejected according to the bootstrapped p-values. The BDS test results indicate that both the GSTUR class models and the AR(p) models could well represent the underlying process of the UK inflation.

[7]In this case, the deterministic structures are GSTUR without constant or trend with lag length specified at 0,1,4 and 5. The p is selected as 4 and 5 in the stationary AR(p) model.

Table 5.13: Small Sample: BDS Test

	m=2		m=3		m=4		m=5	
	BDS-stat	P-value	BDS-stat	P-value	BDS-stat	P-value	BDS-stat	P-value
GS0	.032327	.1238	.042988	.1822	.048184	.2020	.077338	.0876
GS1	.002192	.7000	.012339	.5360	.015155	.5400	.008842	.6080
GS4	-.018759	.5480	-.003520	.8280	-.012236	.9080	-.003557	.7120
GS5	-.034167	.2400	-.052850	.3800	-.041583	.6200	-.104655	.0800
SB	.043056	.0320	.050776	.0880	.060102	.0920	.087916	.0160
RW	.052420	.0000	.061796	.0000	.074011	.0000	.091968	.0000
AR(5)	-.004739	.8600	-.000620	.7400	-.011370	.9000	-.061940	.2400
AR(4)	.002222	.7400	-.009325	.9760	-.019780	.8800	-.076605	.1960

Table 5.14: Forecast of Inflation Rates in Small Sample GSTUR p=1 lag=4 with NO CT

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.2421	3.2512	2.9327	9.8394	-1.0104	9.3728	-57.6152
05Q1	3.17	3.5453	3.4519	3.2298	11.5308	-0.7222	9.3931	-59.1707
05Q2	3.01	2.8288	2.2086	2.6251	5.3872	-1.0143	7.8445	-59.7249
05Q3	2.77	2.5340	2.2455	2.3474	5.3090	-1.0860	7.1594	-60.9709
05Q4	2.38	2.3056	1.5037	2.1885	4.0578	-1.0392	6.4187	-61.9506
06Q1	2.39	2.0845	1.5565	1.9986	3.2559	-1.1018	5.7060	-63.0631
06Q2	2.93	2.2427	1.8924	2.1438	2.7698	-0.7267	5.7828	-63.4988
06Q3	3.44	3.1293	3.0842	3.0280	3.3382	-0.0801	6.9634	-64.4533
06Q4	3.99	3.7835	2.9821	3.6232	3.9126	0.5318	7.8785	-65.5945
07Q1	4.55	4.2063	3.6027	4.0548	4.0159	0.8269	8.7411	-66.8588

Table 5.15: Forecast of Inflation Rates in Small Sample AR(5)

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	MLint
04Q4	3.41	2.8917	3.0426	2.8998	1.2886	0.6047	5.1305	-26.3114
05Q1	3.17	3.1886	3.3225	3.1980	1.2263	0.9690	5.3196	-27.0599
05Q2	3.01	2.9519	2.8729	2.9520	1.1407	0.8079	5.0341	-27.7352
05Q3	2.77	2.7230	2.6637	2.7140	1.0728	0.6631	4.8170	-28.3648
05Q4	2.38	2.5430	2.3091	2.5397	1.0189	0.5537	4.5665	-28.9798
06Q1	2.39	2.2864	2.1462	2.2769	0.9585	0.3299	4.2261	-29.5957
06Q2	2.93	2.3120	2.4004	2.3273	0.9206	0.4143	4.1865	-30.1831
06Q3	3.44	2.8706	2.8437	2.8759	0.9475	0.9269	4.7720	-31.0986
06Q4	3.99	3.3989	3.5266	3.4158	0.9496	1.4353	5.3284	-31.8517
07Q1	4.55	3.8215	3.7729	3.8438	0.9172	1.8920	5.6708	-32.5340

Forecast Results

Following exact the same forecast constructing and forecast evaluating procedures in section 5.4.2, the descriptive statistics and plots of the forecast densities are provided. Density forecasts are evaluated using the PIT approach and point forecasts are evaluated using the MSFE. Forecasting results from other models of interest can be found in Appendix 5.B.

Tables (5.14, 5.15, 5.16, 5.17) provide descriptive statistics of the forecast distributions using the GSTUR forecasting model without a constant or trend at lag 4, AR(5) forecasting model, SB and RW forecasting model.

Figures 5.g(1), 5.h(1), 5.i(1), and 5.j(1) plot the forecast densities from GSTUR

Table 5.16: Forecast of Inflation Rates in Small Sample with a SB model

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	2.7606	2.8833	2.7940	0.9999	0.7232	4.6580	-7.9684
05Q1	3.17	3.0524	2.7180	3.0582	0.9736	1.0459	4.9509	-5.1959
05Q2	3.01	2.8861	3.1170	2.9252	0.8763	0.9804	4.6837	-3.7164
05Q3	2.77	2.7520	2.7208	2.7695	0.8425	0.8488	4.5753	-5.5237
05Q4	2.38	2.5149	2.4170	2.5204	0.7712	0.6960	4.2294	-1.5991
06Q1	2.39	2.1779	2.2308	2.1956	0.7328	0.4291	3.8687	0.3904
06Q2	2.93	2.1565	2.4331	2.1681	0.7118	0.4495	3.7897	1.0419
06Q3	3.44	2.6354	2.5173	2.6422	0.7672	0.8195	4.2967	2.1717
06Q4	3.99	3.0546	3.3315	3.0959	0.8084	1.2142	4.7280	1.5427
07Q1	4.55	3.5592	3.6924	3.5886	0.8063	1.7013	5.2522	3.2170

Table 5.17: Forecast of Inflation Rates in Small Sample with a RW Model

		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.1451	3.2226	3.1420	0.2371	2.2008	4.1109	-22.5721
05Q1	3.17	3.4810	3.5361	3.4828	0.2271	2.5568	4.4162	-23.3848
05Q2	3.01	3.2054	3.3324	3.2045	0.2224	2.2774	4.1262	-24.1493
05Q3	2.77	3.0373	3.0795	3.0407	0.2121	2.1385	3.9318	-24.8684
05Q4	2.38	2.7842	2.7233	2.7796	0.2071	1.8956	3.6782	-25.5950
06Q1	2.39	2.3743	2.3500	2.3750	0.2057	1.4912	3.2556	-26.3884
06Q2	2.93	2.3840	2.3264	2.3798	0.1905	1.5343	3.2477	-27.0319
06Q3	3.44	2.9398	2.9109	2.9399	0.2021	2.0505	3.8161	-27.9311
06Q4	3.99	3.4850	3.4733	3.4821	0.1999	2.6054	4.3591	-28.7960
07Q1	4.55	4.0753	4.0877	4.0747	0.2003	3.2104	4.9480	-29.6969

without constant or trend lag 5, stationary AR(5), SB, and RW forecasting models over 10 forecast period, and Figures 5.g(2), 5.h(2), 5.i(2), and 5.j(2) plot the correlograms of the PITs on the levels and higher powers from the forecast densities accordingly.

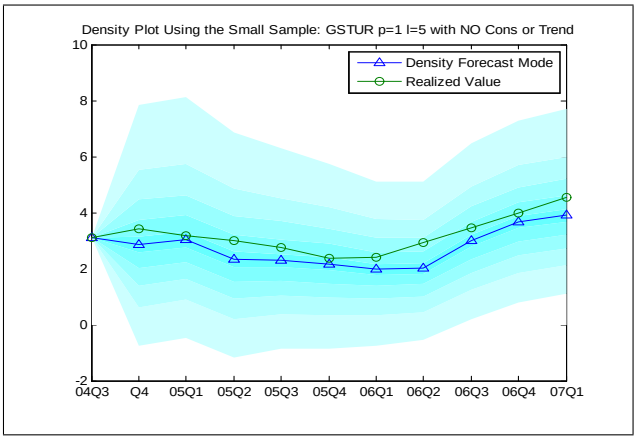


Figure 5.g(1): Small Sample with GSTUR no CT, lag=5

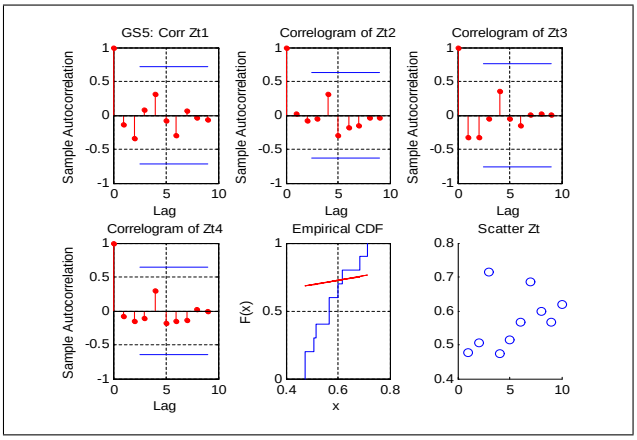


Figure 5.g(2): Correlogram of PITs from GSTURnoCT lag5

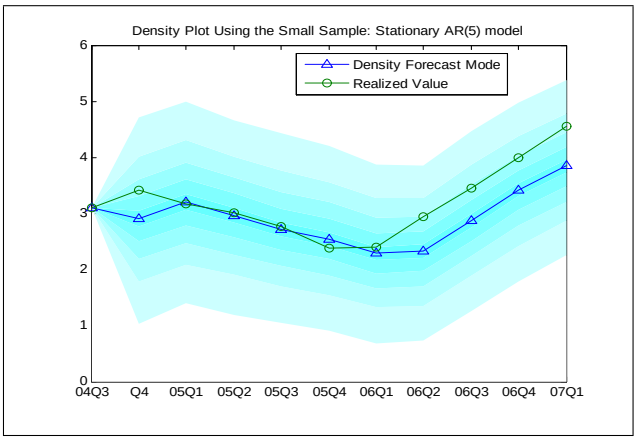


Figure 5.h(1): Small sample Stationary AR(5)

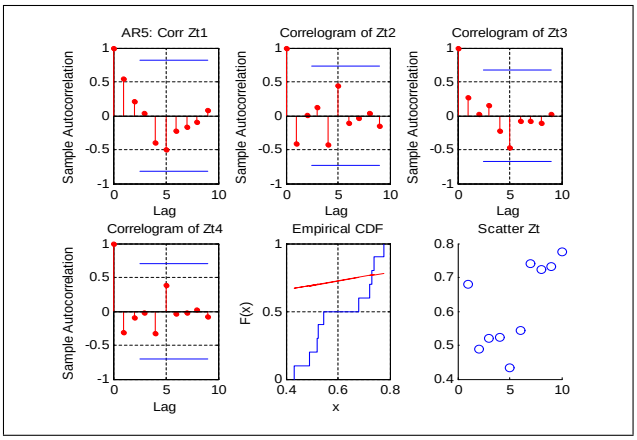


Figure 5.h(2): Correlogram of PITs from stationary AR(5)

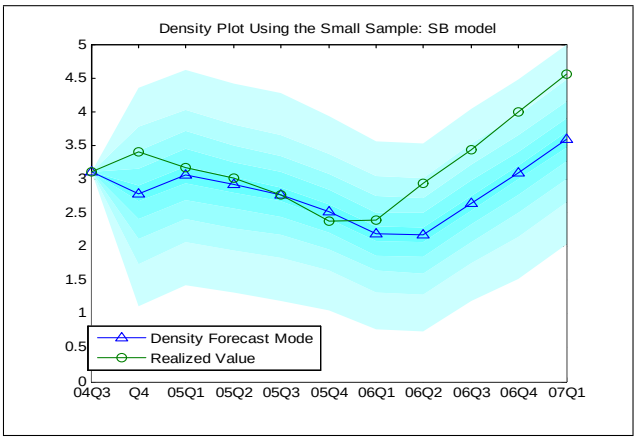


Figure 5.i(1): Small sample SB

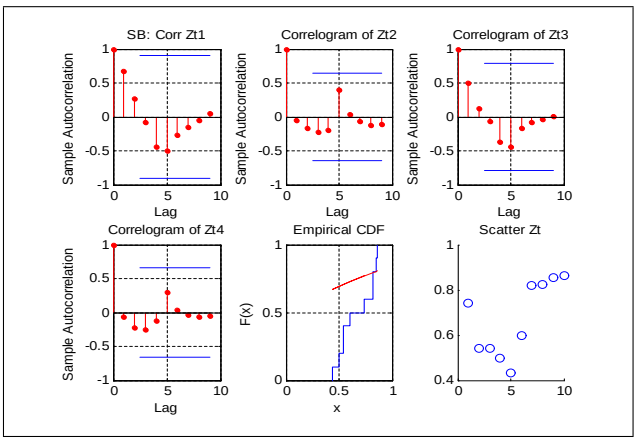


Figure 5.i(2): Correlogram of PITs from SB

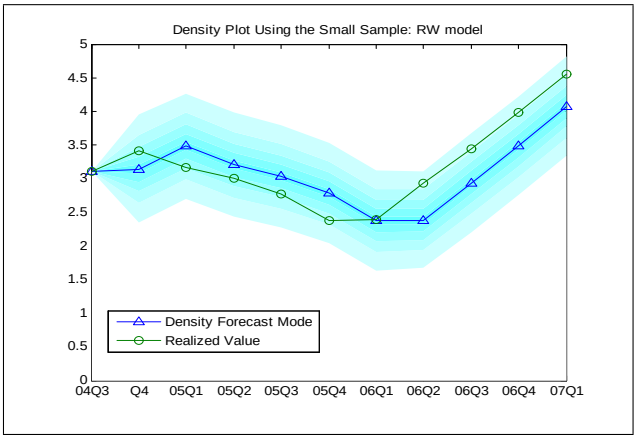


Figure 5.j(2): Small sample RW

Table 5.18: Small Sample: KS Test for PIT i.i.d.U

SB	GSnoCT:$l=0$	GSnoCT:$l=1$	GSnoCT:$l=4$
1	1	1	1
GSnoCT:$l=5$	RW	AR(4)	AR(5)
1	0	1	1

Table 5.19: Small Sample: MSFE of Statistical Forecasting Models

SB	GSnoCT:$l=0$	GSnoCT:$l=1$	GSnoCT:$l=4$
0.389	0.133	0.1021	0.0987
GSnoCT:$l=5$	RW	AR(4)	AR(5)
0.1268	0.1335	0.2199	0.1879

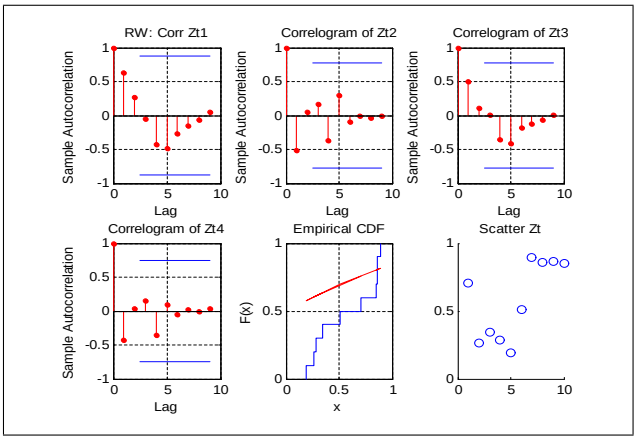


Figure 5.j(2): Correlogram of PITs from RW

Table (5.18) presents the KS test results from selected models, and indicates the null hypothesis that U(0,1)/N(0,1) cannot be rejected by the KS test. Table (5.19) presents the MSFE from the forecasting models.

According to Table (5.19), a GSTUR without a constant or trend, with lag length selected at 4 provides the best point forecast with the smallest value of MSFE amongst all selected models. Although the data series exhibit linearity, the nonlinear GSTUR model could still provide a better point forecast than a stationary linear AR(5) model. This indicates that the GSTUR model we proposed could capture the underlying dynamics of inflation not only when the data exhibits nonlinearity, but also when the data series exhibit linearity. However, using the small sample does not improve on the point forecast accuracy if we compare between Table (5.10) and Table (5.19).

According to the results from the KS test, the null hypothesis is not rejected only in the PITs from the RW forecast density. Looking at Figures 5.g(2), 5.h(2), 5.i(2),

Table 5.20: Small Sample: Simple Averaging Density Forecasts

		Density forecast				95% percentile	
	TrueV	mean	mode	median	var	2.5%	97.5%
04Q4	3.41	2.8917	3.0426	2.8998	1.2886	0.6047	5.1305
05Q1	3.17	3.1886	3.3225	3.1980	1.2263	0.9690	5.3196
05Q2	3.01	2.9519	2.8729	2.9520	1.1407	0.8079	5.0341
05Q3	2.77	2.7230	2.6637	2.7140	1.0728	0.6631	4.8170
05Q4	2.38	2.5430	2.3091	2.5397	1.0189	0.5537	4.5665
06Q1	2.39	2.2864	2.1462	2.2769	0.9585	0.3299	4.2261
06Q2	2.93	2.3120	2.4004	2.3273	0.9206	0.4143	4.1865
06Q3	3.44	2.8706	2.8437	2.8759	0.9475	0.9269	4.7720
06Q4	3.99	3.3989	3.5266	3.4158	0.9496	1.4353	5.3284
07Q1	4.55	3.8215	3.7729	3.8438	0.9172	1.8920	5.6708

and 5.j(2), PITs and higher moments from the GSTUR model display smaller serial correlations. Because the sample size of PITs is as small as 10, conclusive decisions are not made.

Across the GSTUR class of models, a model with lag length 4 provides the best point forecast according to Table (5.19). If we look into Figure 5.g(2) and Table (5.14), the variance of the forecast distribution is very large in the first 2-3 forecast periods. The big variances indicate big forecast uncertainties. However, after the size of sample increases, the variances of the one-step ahead out of sample forecast distributions decrease. Thus, the big forecast uncertainty may be due to the small size of the sample. Compared to results in Appendix, Table for full sample, forecasting inflation rate using GSTUR with noCT, lag=4, where the same model specification was applied to a larger sample, the forecast variances are much smaller.

According to Table (5.16), the SB does not improve on other entertained models, although it receives the highest marginal likelihood throughout the 10 forecast periods. Hence, it may not be appropriate to include the SB model as a constituent forecasting model when combine the forecast via the simple averaging approach.

Forecast Combinations

Performing BMA, because the SB model is the most favoured model according to its marginal likelihoods, the combined forecast density will be dominated by the forecast density of the SB model. Although only RW passed the KS test, for simple averaging, we choose four models, which are GSTUR without constant or trend, lag length specified as 1, 4 and 5, and a stationary AR(5) model.

Table (5.20) presents the descriptive statistics of the combined density via a simple averaging approach. The MSFE over the 10 forecast periods is 0.1708, which is slightly smaller than the MSFE from the stationary AR(5) model.

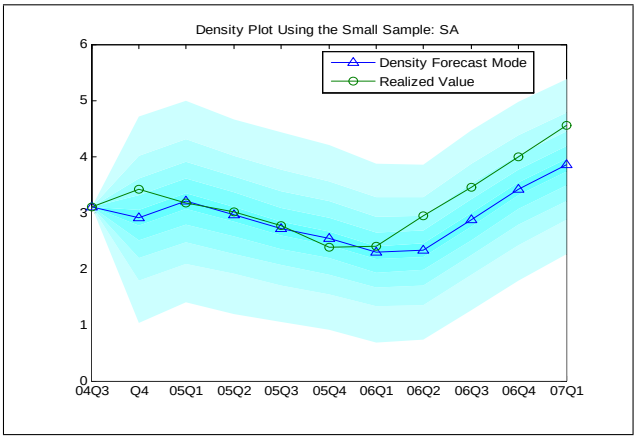


Figure 5.k(1): Combined Density Plot

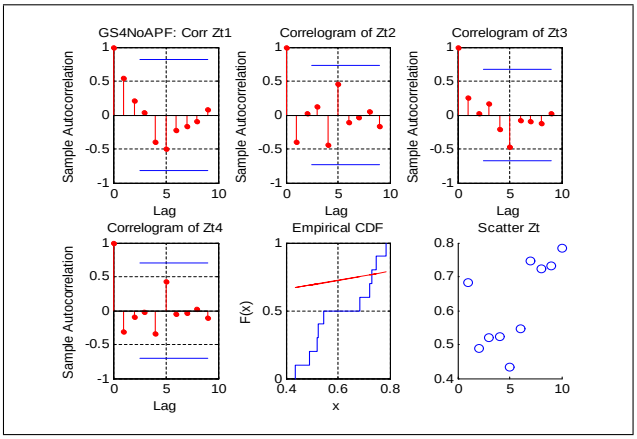


Figure 5.k(2): Correlogram of PITs from Combined Density

According to the KS test, the U(0,1)/N(0,1) in the PITs from the combined density cannot be rejected. The combined density provides very good point forecast from 2005Q1 to 2006Q1. However, after 2006Q2, the combine density exhibits larger forecast errors and the point forecasts are similar to those from a RW forecasting model. In this case, a simple averaging over the four selected models do not improve upon a GSTUR model without constant or trend, with a lag length at 4.

The GSTUR model performs well with both full sample and small sample, and the results are consistent.

From a visual inspection, the full sample of data appears to have more compli-

cated dynamics. Therefore, we expect the proposed GSTUR model to produce better in-sample-fit and better out-of-sample forecast using the full sample. From the out-of-sample point forecast results using the full sample, the GSTUR model indeed outperforms other constituent models in the model space (see the Mean Square Forecast Errors from Table 5.10 on pp.152).

Then, we used a small sample of data (starting from 1999Q4) and applied the same models to forecast the same period (2004:Q4-2007:Q1). From a visual inspection, we may expect the simple linear models to produce better forecast because the small sample has less dynamics and we excluded the possible structural breaks in 1997 by expressly selecting the data period after 1999Q4. By using the Mean Squared Forecast Errors (MSFE), the GSTUR still outperforms the linear AR model and RW model in point forecasting. However, if we look at Figure 5.g(1) and Figure 5.h(1), the forecast density for 2004Q4, the GSTUR noCT forecast density has a much larger forecast error variance than that from the AR(5).

Therefore, if the data appears to be linear and has less dynamics, the GSTUR model may produce smaller MSFE. However, the forecast density may have a large variance because the data is not rich enough for the GSTUR model. The data will be overfitted by a GSTUR model. If one is more interested in the forecast densities, the linear models might be a better choice with the small sample.

5.5 Concluding Remarks

In this chapter, a completed toolbox to calculate substantive quantities, such as model-based point and density forecast, is made available for empirical practice. With the MCMC methods developed in chapter 3 and 4, the density forecasts using statistical forecasting GSTUR and SB models are achieved without extra costs. The point forecasts accuracy are evaluated with Mean Square Forecast Errors and the density forecasts are evaluated with an approach of Probability Integral Transform. Regarding the issues of combining forecasts, we compared a Simple Averaging approach with a Bayesian Model Averaging approach.

Amongst the selected statistical forecasting models in a small model space (24 entertained models), the GSTUR class models are better in providing the point forecasts. Regardless of the size of the samples and the analyzed data period, the GSTUR model outperforms the SB model, stationary AR(p) model and RW model in terms of point forecast. Hence, the nonlinear GSTUR models are not only resilient in respect to the shifts or shocks in the economic system, but are also able to capture the dynamics in

the underlying process of inflation. In this sense, we propose modelling the inflation dynamics using a GSTUR process as it may be able to accommodate breaks in the structure and the changing persistence. Compared with a benchmark RW model, the GSTUR model could provide better point forecasts of inflation.

The combined forecast improves over the linear stationary AR(p) forecasting model. However, the values of MSFE from the combined point forecasts are not smaller than those from a GSTUR model with a constant, using both the BMA and SA approaches. The reason for the unsuccessful combined forecast may be due to identification problems in the forecasting model space. A simple averaging method is superior to a BMA method in the forecast combinations.

A GSTUR model without a constant or trend could detect a significant shift in the persistence. We forecasted the same period from 2004Q4 to 2007Q1 using both the full size of sample and a small size of sample. The GSTUR outperforms the stationary AR(p) models in both cases. If we focus on applying the GSTUR model, simply abandoning part of the data may induce big forecast uncertainties, even though the dynamics of the underlying process might have changed. Therefore, in practice, we may start off using the full sample of data. Then, compare the forecast results from those using a small sample.

In the Bank of England's forecasting procedure, all shocks have to identified, the changes in the variables have to be verified and different forecasting model are reweighted based on the MPC's judgement. Therefore, the fan charts provided by the Bank of England are not only model-based forecast, but also are adjusted according to the MPC's judgement. One question for future research will be, does the Bank of England's fan chart provide enough creditability? More explicitly, should these professional judgements in the MPC be fully trusted, or should more statistical methods be explored to provide forecasts, or should we compromise ourselves by combining the professional's forecasts with the econometric forecasts? Similar questions have been asked in Dowd (2007), Clements (2004) and Cogley et al.(2005). For future research, I will investigate forecasts and evaluate forecasts based on pure statistical models and compare the forecast results with the Bank of England's fan charts.

Appendix 5.A Calculation Issues with AR(p) Model

This Appendix includes the model estimation and marginal likelihood evaluation in the context of a stationary AR(p) model. For more details, refer to Chib (1993).

5.A.1 The Stationary AR(p) Model

y_t is assumed to follow a stationary AR(p) process:

$$y_t = \rho_0 + \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \varepsilon_t \tag{5.7}$$

or can be written as

$$\rho(z)\, y_t = \rho_0 + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. f_N(0, \sigma_\varepsilon^2)$ and $\rho(z) = 1 - \rho_1 z - \cdots - \rho_p z^p$ is a polynomial of order p in the lag operator. To avoid complicates relating to the treatment of initial values, we focus on a stationary AR(p) process and $(y_1, \cdots, y_p)'$ are treated as the initial values. If the series is nonstationary or near so, the treatments of initial values become very important and subtle (see Bauwens, Lubrano and Richard 1999, pp.170-182). Once the AR(p) process is stationary, the treatment of initial conditions will be of less importance. Hence, the characteristic roots of the equation:

$$a^n - \rho_1 a^{n-1} - \cdots - \rho_p = 0$$

all lie within the unit circle. And the characteristic roots of the inverse characteristic equation $\rho(z)$:

$$1 - \rho_1 z - \cdots - \rho_p z^p = 0$$

, should all lie outside of the unit circle[8] to meet the stationary condition. The values of z are the reciprocals of the values of a. We denote Z as a region that $\|z_j\| > 1$ and $\Pr(Z : \|z_j\| > 1)$ is simplified as $\Pr(Z)$. Thus, we assume the prior

$$p\left(\rho_0, \cdots, \rho_p, \sigma_\varepsilon^2\right) = \prod_{i=0}^{p} p\left(\rho_i | \sigma_\varepsilon^2\right) p\left(\sigma_\varepsilon^2\right)$$

is with a Normal-Inverse Gamma form. If we denote $\rho = \left(\rho_0, \cdots, \rho_p\right)'$, the form of the prior suggests that the conditional distribution of ρ given σ_ε^2 should be a multivariate Normal, truncated to the stationary region:

$$p\left(\rho \mid \sigma_\varepsilon^2\right) \propto f_{MN}\left(\underline{\mu}_\rho, \sigma_\varepsilon^2 \underline{V}_\rho\right) 1\left(\|z_j\| > 1\right)$$

where $1(A)$ is the indicator function for the event A, and the marginal distribution for σ_ε^2 is an Inverse-Gamma distribution with the form $f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, where $\underline{\alpha}_\varepsilon$ determines

[8] In other words, greater than one in absolute values.

the shape of the prior and $\underline{\beta}_{\varepsilon}$ determines the scale of the prior. Also, we could assume the prior of ρ, $p(\rho)$ is independent on the prior of σ_{ε}^2, $p(\sigma_{\varepsilon}^2)$. Then the variance covariance matrix in the prior $p(\rho)$ can be expressed as $\underline{V}_{\rho}$.

5.A.2 Model Estimation

Estimating a AR (p) model can also be found in Koop, Poirier and Tobias (2007, chap17). As the initial values are fixed, with a sample of size N, we denote $Y = (y_{p+1}, \cdots, y_{n-1}, y_n)'$, $\rho = (\rho_0, \cdots, \rho_p)$ and

$$X = \begin{bmatrix} 1 & y_p & \cdots & y_1 \\ \vdots & & & \vdots \\ 1 & y_{T-1} & \cdots & y_{T-p} \end{bmatrix}$$

The likelihood function is:

$$p\left(y \mid \rho, \sigma_{\varepsilon}^2\right) = \frac{1}{\left(2\pi\sigma_{\varepsilon}^2\right)^{\frac{N-p}{2}}} \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^2}(Y - X\rho)'(Y - X\rho)\right\}$$

A joint prior is as follows:

$$\begin{aligned} p\left(\rho, \sigma_{\varepsilon}^2\right) \quad \propto \quad & \frac{\left|\sigma_{\varepsilon}^2 \underline{V}_{\rho}\right|^{-1/2}}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^2}\left(\rho - \underline{\mu}_{\rho}\right)' \underline{V}_{\rho}^{-1}\left(\rho - \underline{\mu}_{\rho}\right)\right] \\ & \cdot \frac{1\left(\|z_j\| > 1\right)}{\Pr\left(Z_1\right)} \cdot \frac{1}{\underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}} \Gamma\left(\underline{\alpha}_{\varepsilon}\right)\left(\sigma_{\varepsilon}^2\right)^{\underline{\alpha}_{\varepsilon}+1}} \exp\left(-\frac{1}{\underline{\beta}_{\varepsilon}\sigma_{\varepsilon}^2}\right) \end{aligned}$$

Therefore, the marginal prior for σ_{ε}^2 will be with an Inverse-Gamma form

$$p\left(\sigma_{\varepsilon}^2\right) \propto \frac{1}{\underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}} \Gamma\left(\underline{\alpha}_{\varepsilon}\right)\left(\sigma_{\varepsilon}^2\right)^{\underline{\alpha}_{\varepsilon}+1}} \exp\left(-\frac{1}{\underline{\beta}_{\varepsilon}\sigma_{\varepsilon}^2}\right)$$

and the conditional prior $p(\rho \mid \sigma_{\varepsilon}^2)$ will be:

$$p\left(\rho \mid \sigma_{\varepsilon}^2\right) \propto \frac{\left|\sigma_{\varepsilon}^2 \underline{V}_{\rho}\right|^{-1/2}}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^2}\left(\rho - \underline{\mu}_{\rho}\right)' \underline{V}_{\rho}^{-1}\left(\rho - \underline{\mu}_{\rho}\right)\right] \cdot \frac{1\left(\|z_j\| > 1\right)}{\Pr\left(Z_1\right)}$$

, in which the normalizing constant $\Pr(Z_1)$ is considered to make the prior proper[9].

[9]When $p = 1$, with a univariate normal prior, the normalizing constant will be the c.d.f. within the unit circle $\Phi\left(1, \underline{\mu}_{\rho}, \underline{V}_{\rho}\right) - \Phi\left(-1, \underline{\mu}_{\rho}, \underline{V}_{\rho}\right)$. This normalizing constant makes the truncated prior

Then, according to the Bayes rules, the posterior

$$
\begin{aligned}
p\left(\rho, \sigma_{\varepsilon}^{2} \mid y\right) &\propto p\left(y \mid \rho, \sigma_{\varepsilon}^{2}\right) p\left(\rho, \sigma_{\varepsilon}^{2}\right) \\
&\propto \frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{\frac{N-p}{2}}} \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^{2}}\left(Y-X\rho\right)'\left(Y-X\rho\right)\right\} \\
&\quad \cdot \frac{\left|\sigma_{\varepsilon}^{2}\underline{V}_{\rho}\right|^{-1/2}}{\left(2\pi\right)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left(\rho-\underline{\mu}_{\rho}\right)'\underline{V}_{\rho}^{-1}\left(\rho-\underline{\mu}_{\rho}\right)\right] \\
&\quad \cdot \frac{1\left(\left\|z_{j}\right\|>1\right)}{\Pr\left(Z_{1}\right)} \cdot \frac{1}{\underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}}\Gamma\left(\underline{\alpha}_{\varepsilon}\right)\left(\sigma_{\varepsilon}^{2}\right)^{\underline{\alpha}_{\varepsilon}+1}} \exp\left(-\frac{1}{\underline{\beta}_{\varepsilon}\sigma_{\varepsilon}^{2}}\right) \\
&\propto \frac{\left|\underline{V}_{\rho}\right|^{-1/2} 1\left(\left\|z_{j}\right\|>1\right)}{\left(2\pi\right)^{\frac{N-p+1}{2}} \Pr\left(Z_{1}\right) \cdot \underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}}\Gamma\left(\underline{\alpha}_{\varepsilon}\right)} \cdot \frac{1}{\left(\sigma_{\varepsilon}^{2}\right)^{\frac{N-p}{2}+\frac{1}{2}+\underline{\alpha}_{\varepsilon}+1}} \cdot \\
&\quad \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^{2}}\left[\begin{array}{c}\left(Y-X\rho\right)'\left(Y-X\rho\right) \\ +\left(\rho-\underline{\mu}_{\rho}\right)'\underline{V}_{\rho}^{-1}\left(\rho-\underline{\mu}_{\rho}\right)\end{array}\right]-\frac{1}{\underline{\beta}_{\varepsilon}\sigma_{\varepsilon}^{2}}\right\} \qquad (5.8)
\end{aligned}
$$

The posterior conditionals for ρ is straightforward to derive, which is a truncated multivariate Normal including the normalizing constant:

$$
p\left(\rho \mid y, \sigma_{\varepsilon}^{2}\right) \propto f_{MN}\left(\overline{\mu}_{\rho}, \overline{V}_{\rho}\right) 1\left(\left\|z_{j}\right\|>1\right) \frac{1}{\Pr\left(Z_{2}\right)}
$$

where

$$
\begin{aligned}
\overline{V}_{\rho} &= \left(\underline{V}_{\rho}^{-1}+X'X\right)^{-1} \\
\overline{\mu}_{\rho} &= \overline{V}_{\rho}\left(\underline{V}_{\rho}^{-1}\underline{\mu}_{\rho}+X'Y\right)
\end{aligned}
$$

As the posterior $p\left(\rho \mid y, \sigma_{\varepsilon}^{2}\right)$ is with a truncated form, sampling can be achieved by taking draws from the untruncated distribution, then simply discard the draws which fall outside the stationary region. The integrating constant also can be calculated through the posterior simulation (Koop, 2003, pp.135). $\Pr\left(Z_{2}\right)$ is simply the proportion of draws that are retained at every pass through the Gibbs sampler. The number of rejected draws are calculated before an acceptable one is found, thus the total number of draws is the number of rejected draws plus one. Hence, $\Pr\left(Z_{2}\right)$ can be approximated by one over the total number of draws.

The posterior conditional for σ_{ε}^{2}:

integrates up to 1, then, proper. When $p>1$, with a multivariate normal prior, the integrating constant should be evaluated via simulations.

$$p\left(\sigma_{\varepsilon}^{2} \mid y\right) \propto \frac{1}{\left(\sigma_{\varepsilon}^{2}\right)^{\overline{\alpha}_{\varepsilon}+1}} \exp\left[-\frac{1}{\overline{\beta}_{\varepsilon}\sigma_{\varepsilon}^{2}}\right]$$

where

$$\overline{\alpha}_{\varepsilon} = \underline{\alpha}_{\varepsilon} + \frac{N-p+1}{2}$$

and

$$\overline{\beta}_{\varepsilon} = \left\{ \underline{\beta}_{\varepsilon}^{-1} + \frac{1}{2}\left[\begin{array}{c} (Y-X\rho)'(Y-X\rho) \\ +\left(\rho-\underline{\mu}_{\rho}\right)' \underline{V}_{\rho}^{-1}\left(\rho-\underline{\mu}_{\rho}\right) \end{array} \right] \right\}^{-1}$$

5.A.3 The Marginal likelihood Evaluation

To approximate the marginal likelihood, we apply the numerical integration method via the MCMC outputs. The conventional Gelfand-Dey method can be implemented and the algorithm is listed as the following:

According $E\left[\frac{f(\theta)}{p(\theta|M_j)p(y|\theta,M_j)} \mid y, M_j\right] = \frac{1}{p(y|M_j)}$, use the posterior draws of $\rho^{*}, \sigma_{\varepsilon}^{2*}$

1. obtain $\widehat{\theta}$ and $\widehat{\sum}$, estimates of $E\left(\theta \mid y, M_j\right)$ and $var\left(\theta \mid y, M_j\right)$ obtained from the posterior simulator. The truncated region is

$$\widehat{\Theta} = \left\{\theta : \left(\theta-\widehat{\theta}\right)' \widehat{\sum}^{-1} \left(\theta-\widehat{\theta}\right) \leq \chi_{1-p}^{2}(k)\right\}$$

2. $f(\theta)$ be the multivariate Normal density truncated to the region $\widehat{\Theta}$,

$$f(\theta) = \frac{1}{p(2\pi)^{\frac{k}{2}}} \left|\widehat{\sum}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta-\widehat{\theta}\right)' \widehat{\sum}^{-1} \left(\theta-\widehat{\theta}\right)\right] 1\left(\theta \in \widehat{\Theta}\right)$$

   if θ^{*} is within $\widehat{\Theta}$, $f(\theta) = f(\theta^{*})$, otherwise 0

3. Evaluate $p\left(y|\rho^{*}, \sigma_{\varepsilon}^{2*}\right)$, and $p\left(\rho^{*}, \sigma_{\varepsilon}^{2*}\right)$

4. $p\left(y \mid M_j\right) \propto E\left[\frac{f(\theta)}{p(\rho^{*},\sigma_{\varepsilon}^{2*}|M_j)p(y|\rho^{*},\sigma_{\varepsilon}^{2*},M_j)} \mid y, M_j\right]^{-1}$

$p\left(\theta \mid M_j\right)$ are evaluated at posterior draws $p\left(\rho^{*}, \sigma_{\varepsilon}^{2*}\right)$

$$\begin{aligned} p\left(\rho^{*}, \sigma_{\varepsilon}^{2*}\right) \quad \propto \quad & \frac{\left|\sigma_{\varepsilon}^{2*}\underline{V}_{\rho}\right|^{-1/2} 1\left(\|z_j\| > 1\right)}{\Pr(Z)(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2*}}\left(\rho^{*}-\underline{\mu}_{\rho}\right)' \underline{V}_{\rho}^{-1}\left(\rho^{*}-\underline{\mu}_{\rho}\right)\right] \\ & \cdot \frac{\underline{\beta}^{\underline{\alpha}}}{\Gamma(\underline{\alpha})\left(\sigma_{\varepsilon}^{2*}\right)^{\underline{\alpha}+1}} \exp\left(-\frac{\underline{\beta}}{\sigma_{\varepsilon}^{2*}}\right) \end{aligned}$$

where $\Pr(Z)$ can be evaluated by simulating ρ from the prior $f_{MN}\left(\underline{\mu}_{\rho}, \underline{V}_{\rho}\right)$ and calculating the proportion of draws from the prior are retained. The likelihood function evaluated at $\rho^*, \sigma_{\varepsilon}^{2*}$ are calculated as the following:

$$p\left(y \mid \rho^*, \sigma_{\varepsilon}^{2*}\right) = \frac{1}{\left(2\pi\sigma_{\varepsilon}^{2*}\right)^{\frac{N-p}{2}}} \exp\left\{-\frac{1}{2\sigma_{\varepsilon}^{2*}}\left(Y - X\rho^*\right)'\left(Y - X\rho^*\right)\right\}$$

Appendix 5.B Forecast Results from Models of interest

5.B.1 Full Sample

Density Forecast Result Tables

Full Sample: Forecast of Inflation Rates GSTUR with p=1, lag=0, with Constant								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.1416	3.0931	3.1466	3.1069	-0.3694	6.5619	-348.8857
05Q1	3.17	3.4494	3.6887	3.4476	3.1081	0.0095	6.9417	-349.9585
05Q2	3.01	3.2326	3.1062	3.2277	3.1210	-0.2201	6.7378	-350.8349
05Q3	2.77	3.0661	3.3494	3.0868	3.0873	-0.4714	6.4582	-352.0960
05Q4	2.38	2.8195	3.0822	2.8413	3.0579	-0.6264	6.1673	-352.6958
06Q1	2.39	2.4513	2.2568	2.4459	3.0486	-0.9999	5.8325	-353.9378
06Q2	2.93	2.4616	2.9798	2.4726	3.1161	-1.0210	5.8869	-354.2599
06Q3	3.44	2.9847	2.4275	2.9992	3.0790	-0.5340	6.4082	-356.1762
06Q4	3.99	3.4772	3.6528	3.4689	3.0142	0.0096	6.9520	-356.8188
07Q1	4.55	4.0109	4.0818	4.0050	3.0120	0.5090	7.4001	-357.6725

Full Sample:Forecast Inflation Rates GSTUR with p=1, lag=0 with Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.1927	2.6195	3.2035	6.3291	-1.8986	8.1137	-348.5880
05Q1	3.17	3.4739	3.5152	3.4792	6.4699	-1.5652	8.5183	-349.7647
05Q2	3.01	3.2459	2.8321	3.2487	6.3051	-1.7270	8.1479	-350.2897
05Q3	2.77	3.0897	2.4782	3.1020	6.4268	-2.0293	7.9807	-351.2280
05Q4	2.38	2.8643	3.1963	2.9056	6.3320	-2.2041	7.8283	-352.0063
06Q1	2.39	2.4495	2.2349	2.4290	6.2749	-2.5489	7.3705	-353.1747
06Q2	2.93	2.4513	2.5815	2.4753	6.4102	-2.6578	7.4518	-354.9464
06Q3	3.44	3.0136	2.9060	3.0265	6.3879	-1.9772	7.9925	-355.7706
06Q4	3.99	3.5001	2.7491	3.4935	5.8777	-1.2791	8.2438	-356.1319
07Q1	4.55	4.0547	4.4979	4.0482	6.0959	-0.8758	8.8955	-357.3092

Full Sample: Forecast of Inflation Rates GSTUR p=1, lag=0 with Cons and Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	MLint
04Q4	3.41	3.1276	3.1192	3.0910	6.7827	-2.0966	8.2492	-367.6159
05Q1	3.17	3.4174	2.8601	3.3946	6.7468	-1.7146	8.4842	-368.4840
05Q2	3.01	3.1794	3.8001	3.2160	6.5116	-1.8572	8.1637	-369.4559
05Q3	2.77	3.0216	2.4982	3.0254	6.7716	-2.1709	8.2006	-370.0992
05Q4	2.38	2.7725	3.0494	2.7775	6.8604	-2.4524	7.9665	-371.6082
06Q1	2.39	2.4082	2.4187	2.4002	6.5073	-2.5918	7.3563	-372.1059
06Q2	2.93	2.4263	2.1830	2.4010	6.4584	-2.5352	7.4265	-373.4161
06Q3	3.44	2.9365	3.3094	2.9451	6.5761	-2.1579	7.9423	-374.9165
06Q4	3.99	3.4445	3.8689	3.4756	6.4759	-1.7347	8.4724	-375.8846
07Q1	4.55	3.9732	4.1671	3.9832	6.3893	-1.0754	8.9924	-376.6038

Full Sample: Forecast of Inflation Rates GSTUR p=1, lag=0 with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.0501	2.9367	3.0323	1.8942	0.3434	5.7931	-364.5002
05Q1	3.17	3.3504	3.4243	3.3507	1.9046	0.6376	6.0509	-365.6233
05Q2	3.01	3.1280	3.0123	3.1223	1.8522	0.4832	5.8181	-366.5946
05Q3	2.77	2.9979	3.3141	3.0251	1.8720	0.2663	5.6958	-368.0697
05Q4	2.38	2.7226	2.4062	2.7042	1.7809	0.1273	5.3689	-368.5695
06Q1	2.39	2.3582	2.2913	2.3581	1.7711	-0.2439	5.0075	-369.9566
06Q2	2.93	2.3449	2.7974	2.3463	1.7671	-0.2699	4.9268	-370.6569
06Q3	3.44	2.8977	2.8681	2.8964	1.7900	0.2911	5.4819	-371.9259
06Q4	3.99	3.3907	3.1326	3.3797	1.8285	0.7861	6.0477	-373.4739
07Q1	4.55	3.9575	3.7339	3.9612	1.8349	1.2657	6.6488	-374.1413

Full Sample: Forecast of Inflation Rates with GSTUR p=1, lag=1 with Cons								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3669	3.5560	3.3946	2.6836	0.0856	6.5423	-330.3961
05Q1	3.17	3.5964	3.5657	3.6040	2.6262	0.3367	6.7344	-331.2558
05Q2	3.01	3.1234	3.3392	3.1450	2.6426	-0.1558	6.2874	-331.8223
05Q3	2.77	2.9939	2.9635	2.9889	2.6278	-0.2243	6.1696	-332.9836
05Q4	2.38	2.7416	2.5298	2.7395	2.6580	-0.5783	5.9106	-334.2732
06Q1	2.39	2.2863	2.4017	2.3100	2.5746	-0.9626	5.3849	-334.2490
06Q2	2.93	2.4755	2.2596	2.4746	2.6631	-0.7798	5.6097	-335.2896
06Q3	3.44	3.2526	3.2921	3.2649	2.6012	0.0319	6.4277	-336.0465
06Q4	3.99	3.7173	3.8278	3.7349	2.5192	0.5351	6.8185	-337.0761
07Q1	4.55	4.2820	4.0507	4.2768	2.4674	1.2522	7.3991	-338.3455

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=1, with Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3499	4.0268	3.3962	5.7172	-1.4091	7.9304	-332.5677
05Q1	3.17	3.6294	4.1553	3.6533	5.7366	-1.1031	8.3026	-333.2003
05Q2	3.01	3.1398	2.4870	3.1445	5.7159	-1.6746	7.8167	-334.6216
05Q3	2.77	3.0324	2.9776	3.0373	5.8150	-1.8067	7.7547	-334.7940
05Q4	2.38	2.7647	2.6680	2.7907	5.6339	-2.0253	7.4453	-335.9560
06Q1	2.39	2.2863	2.1702	2.3156	5.7425	-2.5429	6.9328	-337.1164
06Q2	2.93	2.4949	2.5094	2.5074	5.5695	-2.2781	7.0381	-338.0164
06Q3	3.44	3.2180	2.4571	3.2548	5.6572	-1.6172	7.7694	-338.7976
06Q4	3.99	3.7460	3.7939	3.7671	5.3429	-0.8484	8.1775	-339.5560
07Q1	4.55	4.2991	4.2544	4.3078	5.1882	-0.1805	8.7302	-340.3982

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=1Cons and Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.2733	3.9603	3.2780	5.6915	-1.2958	7.9910	-348.5737
05Q1	3.17	3.5399	2.8210	3.5050	6.0861	-1.3948	8.4454	-349.5522
05Q2	3.01	3.0597	2.5411	3.0370	5.9531	-1.7049	7.9535	-350.0263
05Q3	2.77	2.9538	2.4060	2.9583	5.7493	-1.7789	7.6593	-350.8430
05Q4	2.38	2.6905	2.6916	2.6508	5.6494	-2.0419	7.4461	-351.7856
06Q1	2.39	2.2374	2.6373	2.2368	5.7667	-2.4781	6.9478	-352.7526
06Q2	2.93	2.4133	2.4589	2.4236	5.7134	-2.4023	7.1238	-353.1802
06Q3	3.44	3.1741	2.5746	3.1564	5.6616	-1.5604	7.9037	-354.4472
06Q4	3.99	3.6678	3.8836	3.6768	5.4566	-0.9566	8.2881	-354.7396
07Q1	4.55	4.2056	4.1774	4.2188	5.5450	-0.4271	8.9093	-356.2698

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=1, with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.1908	3.2993	3.1936	1.5972	0.6897	5.6739	-351.5333
05Q1	3.17	3.5252	3.6034	3.5288	1.6577	0.9901	6.0634	-352.5781
05Q2	3.01	3.0346	3.0961	3.0370	1.6341	0.5397	5.5223	-353.7911
05Q3	2.77	2.8855	3.0605	2.8898	1.5499	0.3933	5.3342	-354.1335
05Q4	2.38	2.6293	2.8118	2.6303	1.5575	0.2153	5.1142	-355.5657
06Q1	2.39	2.2126	1.9563	2.2064	1.5106	-0.2325	4.6522	-355.8802
06Q2	2.93	2.3599	2.3714	2.3584	1.5341	-0.0446	4.7808	-356.9000
06Q3	3.44	3.1359	3.3721	3.1456	1.5378	0.6795	5.5634	-357.9808
06Q4	3.99	3.6364	3.5904	3.6406	1.5863	1.1156	6.1068	-358.3212
07Q1	4.55	4.1965	4.3760	4.1928	1.6316	1.7010	6.7073	-359.5996

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=4, with Cons								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4246	3.5251	3.4444	2.4078	0.3558	6.4762	-337.1829
05Q1	3.17	3.6697	3.9774	3.6972	2.4498	0.5482	6.6826	-338.0330
05Q2	3.01	3.1411	3.0527	3.1462	2.3080	0.1702	6.0945	-338.9188
05Q3	2.77	2.9290	2.6508	2.9385	2.4217	-0.1156	5.9282	-340.4070
05Q4	2.38	2.6680	2.9666	2.6918	2.3496	-0.3979	5.5968	-339.9444
06Q1	2.39	2.3609	2.4662	2.3657	2.3942	-0.6685	5.3553	-341.1014
06Q2	2.93	2.5233	2.9071	2.5348	2.4249	-0.5743	5.5269	-342.0057
06Q3	3.44	3.2980	3.1658	3.3065	2.3239	0.2953	6.2502	-342.3122
06Q4	3.99	3.8661	4.0003	3.8750	2.2478	0.9167	6.8703	-343.9123
07Q1	4.55	4.3025	4.4270	4.3171	2.2155	1.3998	7.2242	-344.1763

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=4, with Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4484	3.1324	3.4454	5.1634	-1.1214	7.8862	-337.1736
05Q1	3.17	3.6843	3.4623	3.6897	5.0771	-0.7610	8.1274	-337.6904
05Q2	3.01	3.1441	3.1896	3.1395	5.1274	-1.3835	7.5986	-338.9712
05Q3	2.77	2.9371	3.0097	2.9655	4.9814	-1.5315	7.3231	-339.3407
05Q4	2.38	2.6471	2.8581	2.6315	4.9217	-1.7627	7.0215	-340.5639
06Q1	2.39	2.3742	2.5871	2.3785	4.8407	-2.0332	6.6634	-341.4422
06Q2	2.93	2.5329	3.0343	2.5824	4.8934	-1.8710	6.7830	-341.7596
06Q3	3.44	3.3101	3.1570	3.3281	4.8800	-1.1315	7.6321	-342.6748
06Q4	3.99	3.8847	4.0396	3.8921	4.8150	-0.4062	8.2071	-343.4323
07Q1	4.55	4.3520	4.1209	4.3551	4.6295	0.1378	8.5813	-344.6385

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=4 Cons and Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3387	3.5310	3.3712	5.1230	-1.1690	7.8461	-355.9397
05Q1	3.17	3.6067	3.5636	3.6087	5.1585	-0.8171	8.0385	-356.6626
05Q2	3.01	3.0810	2.9086	3.0582	4.9850	-1.4217	7.4606	-357.9429
05Q3	2.77	2.8580	3.2228	2.8815	5.1411	-1.5600	7.4100	-358.6373
05Q4	2.38	2.5606	2.7945	2.5547	4.9680	-1.7963	6.9120	-359.0501
06Q1	2.39	2.3053	2.2460	2.3004	4.7988	-2.0436	6.5475	-359.3868
06Q2	2.93	2.4151	2.8898	2.4463	4.9657	-1.9798	6.7130	-360.9545
06Q3	3.44	3.1970	3.0150	3.2055	4.9833	-1.2910	7.5669	-361.2716
06Q4	3.99	3.8179	4.0617	3.8409	4.7372	-0.4550	8.1128	-363.4520
07Q1	4.55	4.2383	4.4040	4.2327	4.8757	-0.0533	8.6068	-362.5645

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=4, with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.2739	3.3030	3.2731	1.4501	0.9191	5.6326	-356.0343
05Q1	3.17	3.5239	3.7890	3.5342	1.4178	1.1737	5.8464	-355.9640
05Q2	3.01	3.0118	3.0118	3.0266	1.4353	0.6644	5.3598	-357.1306
05Q3	2.77	2.7696	2.5104	2.7528	1.3939	0.5035	5.1119	-357.8148
05Q4	2.38	2.4993	2.4977	2.5030	1.3949	0.1935	4.7981	-359.2610
06Q1	2.39	2.2635	2.2974	2.2701	1.3344	0.0002	4.4873	-359.7945
06Q2	2.93	2.3523	2.3984	2.3544	1.3059	0.1292	4.5840	-360.4440
06Q3	3.44	3.1735	3.1393	3.1828	1.3583	0.8705	5.4205	-361.4529
06Q4	3.99	3.7707	3.7833	3.7744	1.3976	1.4699	6.1166	-361.9854
07Q1	4.55	4.1975	4.1603	4.1798	1.3977	1.8751	6.5743	-362.5099

Full Sample: Forecast of Inflation Rates with AR(7)							
MSFE		Density forecast				95% percentile	
0.0888	TrueV	mean	mode	median	var	2.5%	97.5%
04Q4	3.41	3.2505	3.0321	3.2268	1.4375	0.8791	5.7085
05Q1	3.17	3.4697	3.8320	3.4911	1.4231	1.1310	5.7062
05Q2	3.01	2.8932	2.4332	2.8821	1.4128	0.5309	5.2125
05Q3	2.77	2.7649	2.8632	2.7649	1.3696	0.4414	5.1134
05Q4	2.38	2.7267	2.8420	2.7258	1.3604	0.5030	5.0368
06Q1	2.39	2.2033	2.2349	2.2077	1.2934	-0.0355	4.4034
06Q2	2.93	2.3797	2.4818	2.3928	1.3069	0.1615	4.6252
06Q3	3.44	3.3137	3.4479	3.2975	1.3177	1.1724	5.5650
06Q4	3.99	3.6527	3.8045	3.6753	1.3879	1.2889	5.8975
07Q1	4.55	4.0400	3.7617	4.0215	1.3566	1.8189	6.3205

Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=5, with Cons								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4813	3.5531	3.4863	2.3044	0.5410	6.4918	-339.6490
05Q1	3.17	3.6422	4.1422	3.6691	2.2612	0.6364	6.5419	-340.3256
05Q2	3.01	3.0789	3.1601	3.0734	2.3020	0.0966	6.0231	-340.9300
05Q3	2.77	2.9213	2.7538	2.9264	2.2346	-0.0237	5.8693	-341.0559
05Q4	2.38	2.6972	2.6788	2.6929	2.3007	-0.2725	5.6798	-341.7964
06Q1	2.39	2.4515	2.7715	2.4814	2.2770	-0.5816	5.3848	-342.9015
06Q2	2.93	2.4788	2.8043	2.5006	2.2833	-0.5561	5.4502	-344.0708
06Q3	3.44	3.3530	3.6669	3.3649	2.2519	0.3855	6.2574	-344.4316
06Q4	3.99	3.8945	4.1122	3.9019	2.2426	0.9175	6.7979	-344.2936
07Q1	4.55	4.2346	4.2865	4.2489	2.1838	1.2899	7.0804	-345.8233

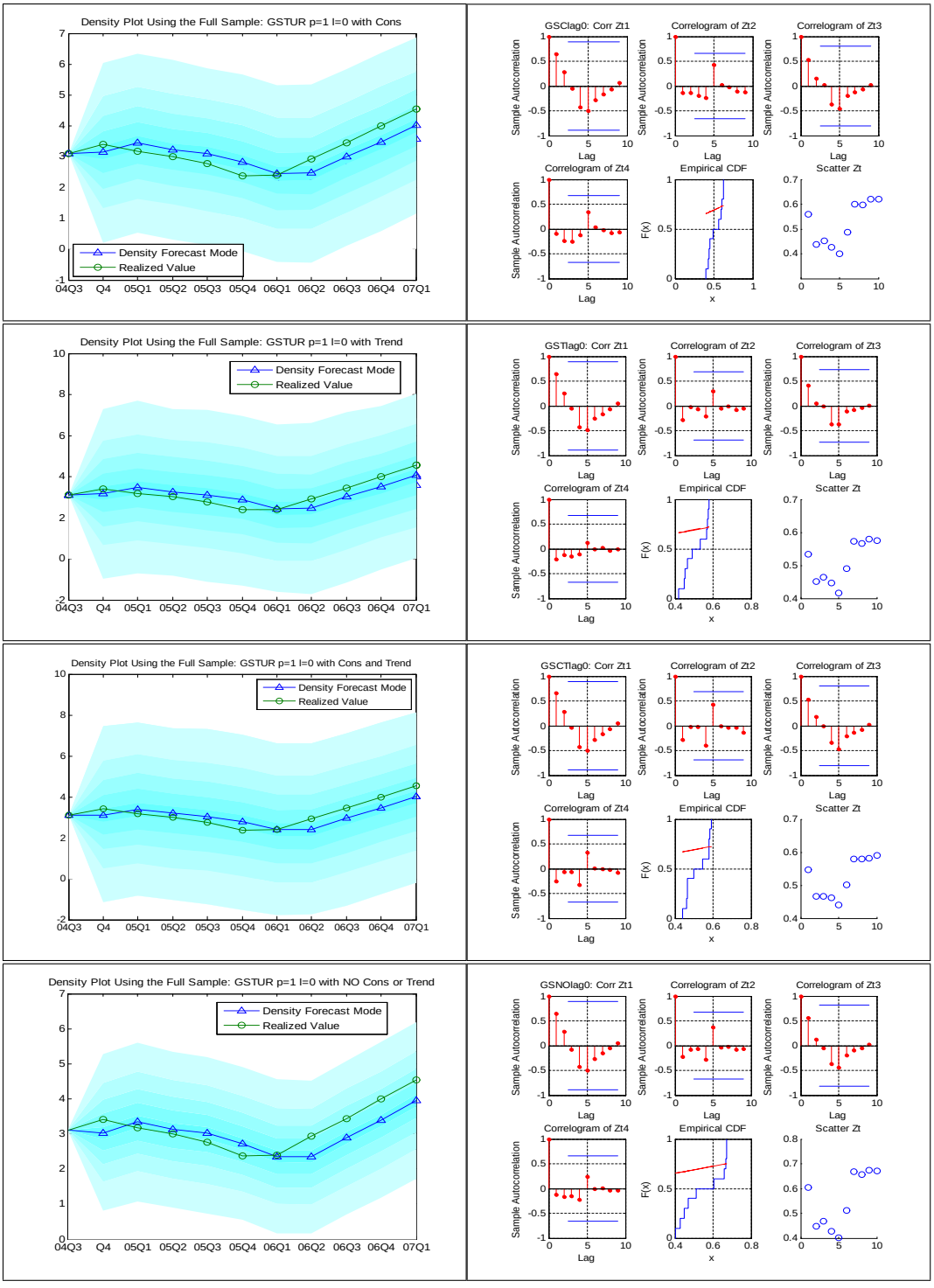
Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=5, with Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4998	3.6374	3.5271	4.9269	-0.8714	7.9112	-339.8521
05Q1	3.17	3.6778	3.3278	3.6908	4.6887	-0.6669	7.8451	-340.8652
05Q2	3.01	3.0902	2.5748	3.0830	4.7706	-1.1902	7.3598	-340.6408
05Q3	2.77	2.9100	3.1869	2.9463	4.8963	-1.5211	7.2481	-341.6210
05Q4	2.38	2.6934	3.0677	2.7191	4.8090	-1.6718	7.0057	-341.6060
06Q1	2.39	2.4564	2.3718	2.4649	4.7271	-1.9418	6.6744	-342.9425
06Q2	2.93	2.4992	2.3903	2.5234	4.7060	-1.8871	6.6501	-343.6233
06Q3	3.44	3.3599	3.2876	3.3802	4.6253	-0.8062	7.5389	-345.1264
06Q4	3.99	3.9141	3.8156	3.9235	4.6944	-0.4642	8.1708	-345.0832
07Q1	4.55	4.2712	4.3773	4.2606	4.7487	-0.0578	8.5673	-345.6120

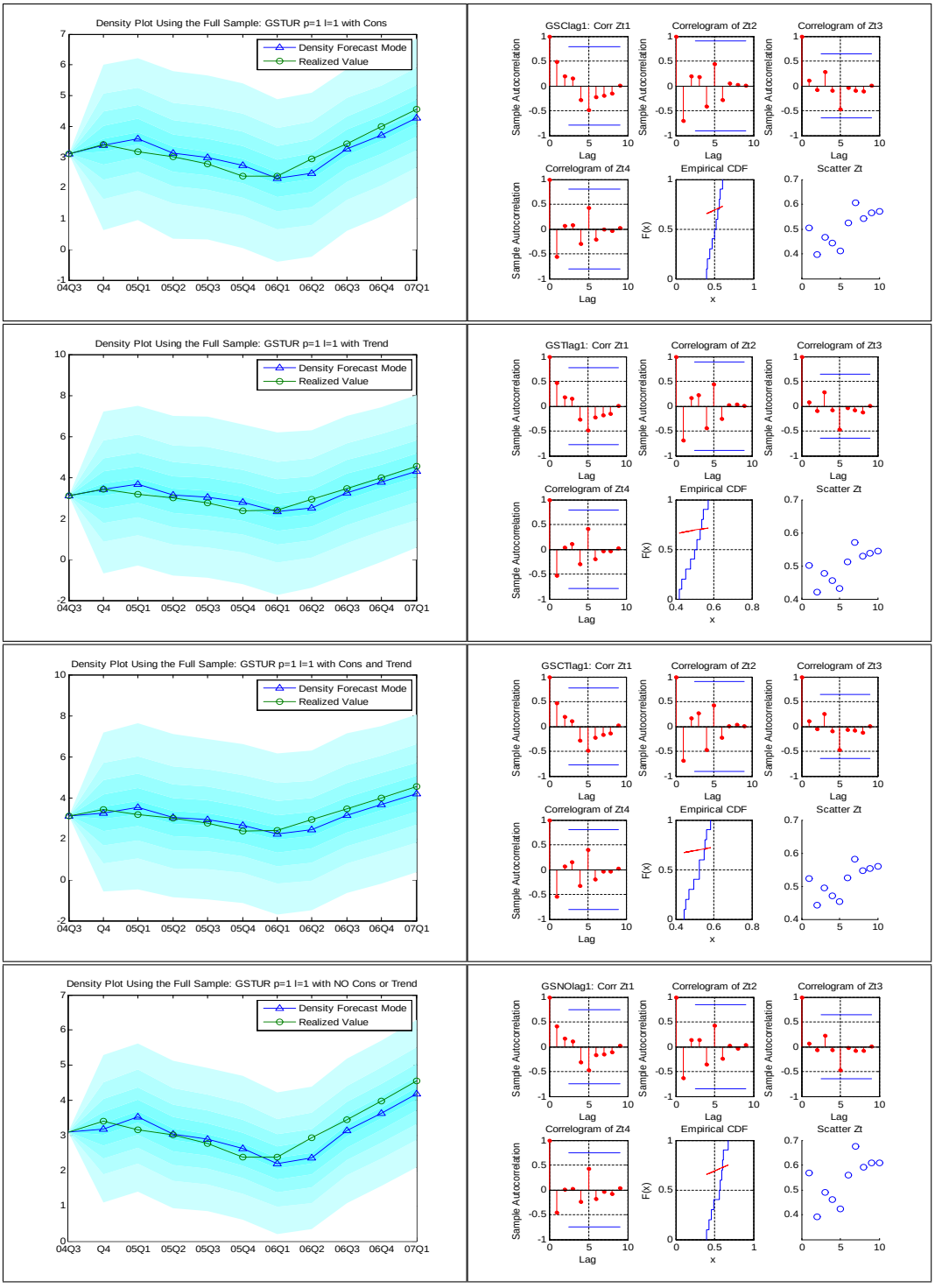
Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=5 Cons and Trend								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.4139	3.7713	3.4329	4.9980	-0.9802	7.9123	-358.0779
05Q1	3.17	3.5588	3.5226	3.5648	4.9760	-0.7988	7.9624	-358.2893
05Q2	3.01	3.0094	2.7436	3.0005	5.1535	-1.4473	7.5002	-359.0244
05Q3	2.77	2.8574	2.6140	2.8344	4.9695	-1.5366	7.3444	-360.2748
05Q4	2.38	2.6334	3.0931	2.6309	4.9796	-1.8119	7.0807	-359.8108
06Q1	2.39	2.3657	2.0941	2.3618	4.6986	-1.9060	6.6364	-360.7820
06Q2	2.93	2.4074	2.2460	2.4018	4.8282	-1.8854	6.7862	-362.6678
06Q3	3.44	3.2782	2.8648	3.2900	4.7928	-1.0591	7.5846	-363.4120
06Q4	3.99	3.8376	4.2998	3.8625	4.6854	-0.5006	8.0811	-363.7807
07Q1	4.55	4.1715	3.8380	4.1679	4.7842	-0.0483	8.4909	-364.6154

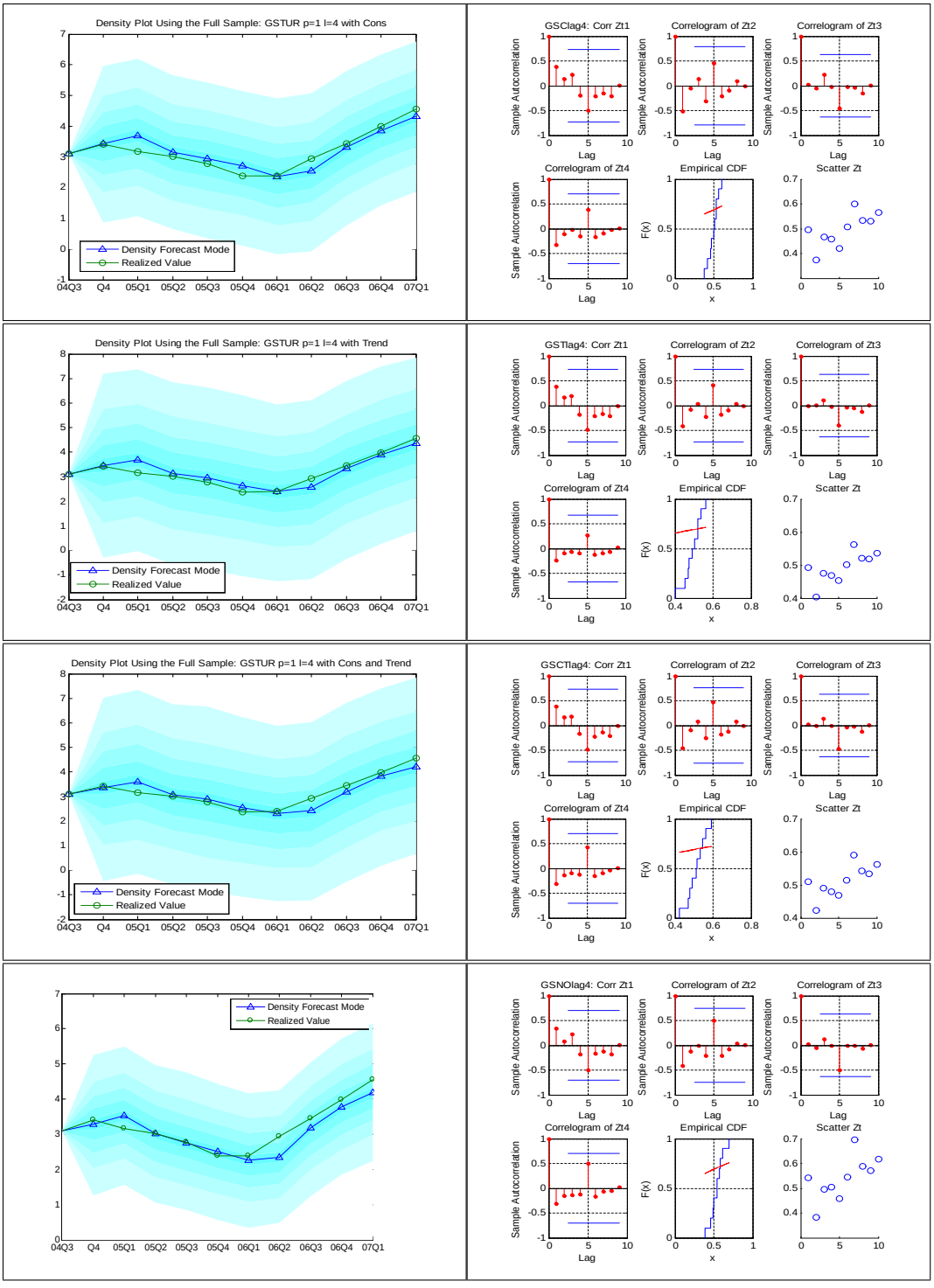
Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=5, with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.3353	3.4371	3.3351	1.3533	1.0895	5.6039	-358.1377
05Q1	3.17	3.5159	3.6377	3.5119	1.3742	1.2810	5.8329	-359.0214
05Q2	3.01	2.8951	3.0150	2.8914	1.3467	0.5930	5.1639	-359.7913
05Q3	2.77	2.7536	2.5226	2.7367	1.3262	0.4931	5.0391	-360.5057
05Q4	2.38	2.5437	2.4817	2.5450	1.2830	0.3271	4.8030	-361.4968
06Q1	2.39	2.3131	2.0932	2.3039	1.2507	0.1329	4.5144	-362.0681
06Q2	2.93	2.3314	2.4389	2.3315	1.2810	0.1523	4.6215	-362.8086
06Q3	3.44	3.2222	3.4040	3.2333	1.2790	0.9589	5.4293	-363.5650
06Q4	3.99	3.7728	3.6984	3.7672	1.3353	1.4984	6.0492	-363.9624
07Q1	4.55	4.1272	4.3988	4.1421	1.3552	1.8448	6.4750	-365.0569

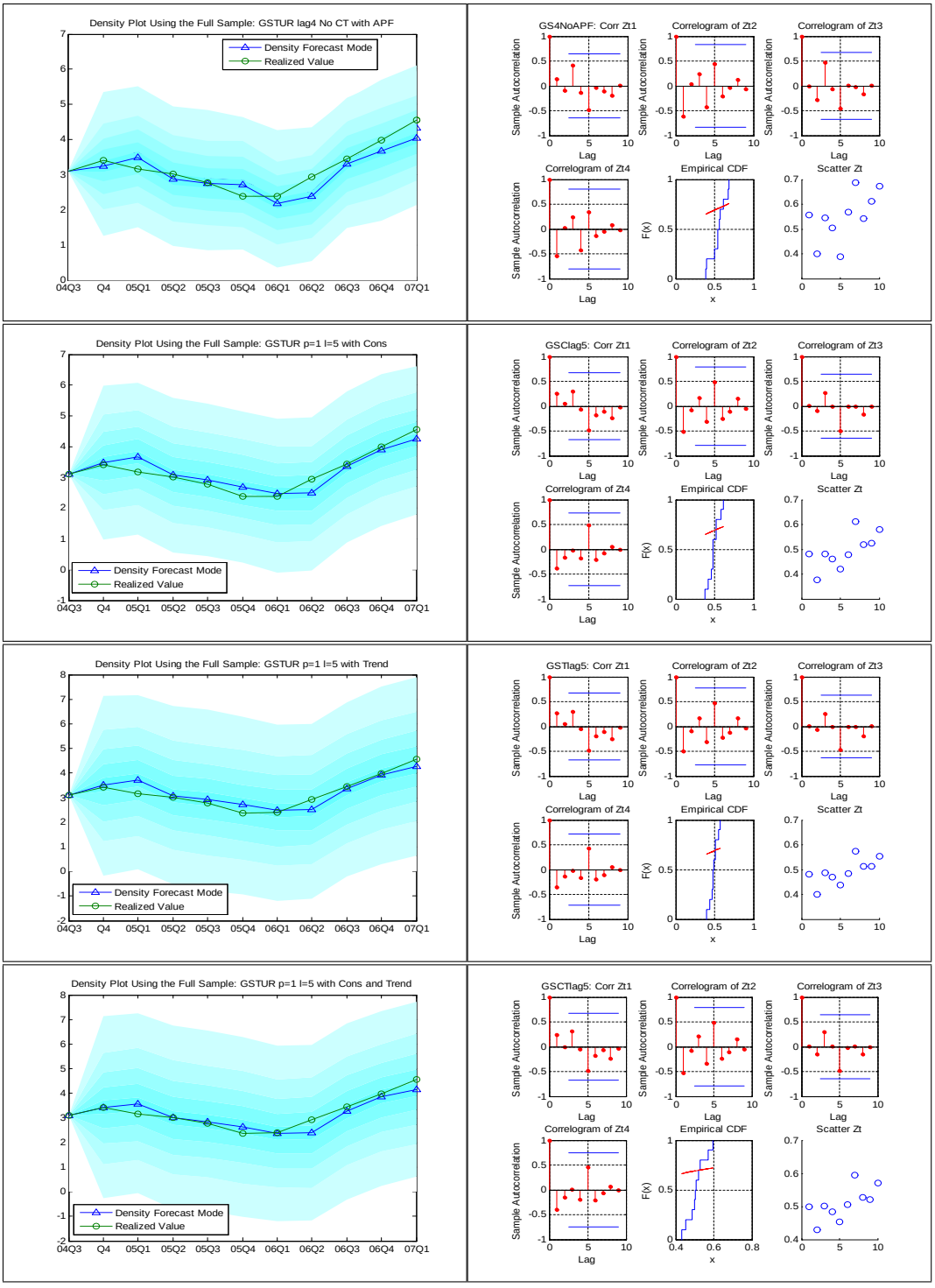
Density Forecast Plots and PIT Evaluation

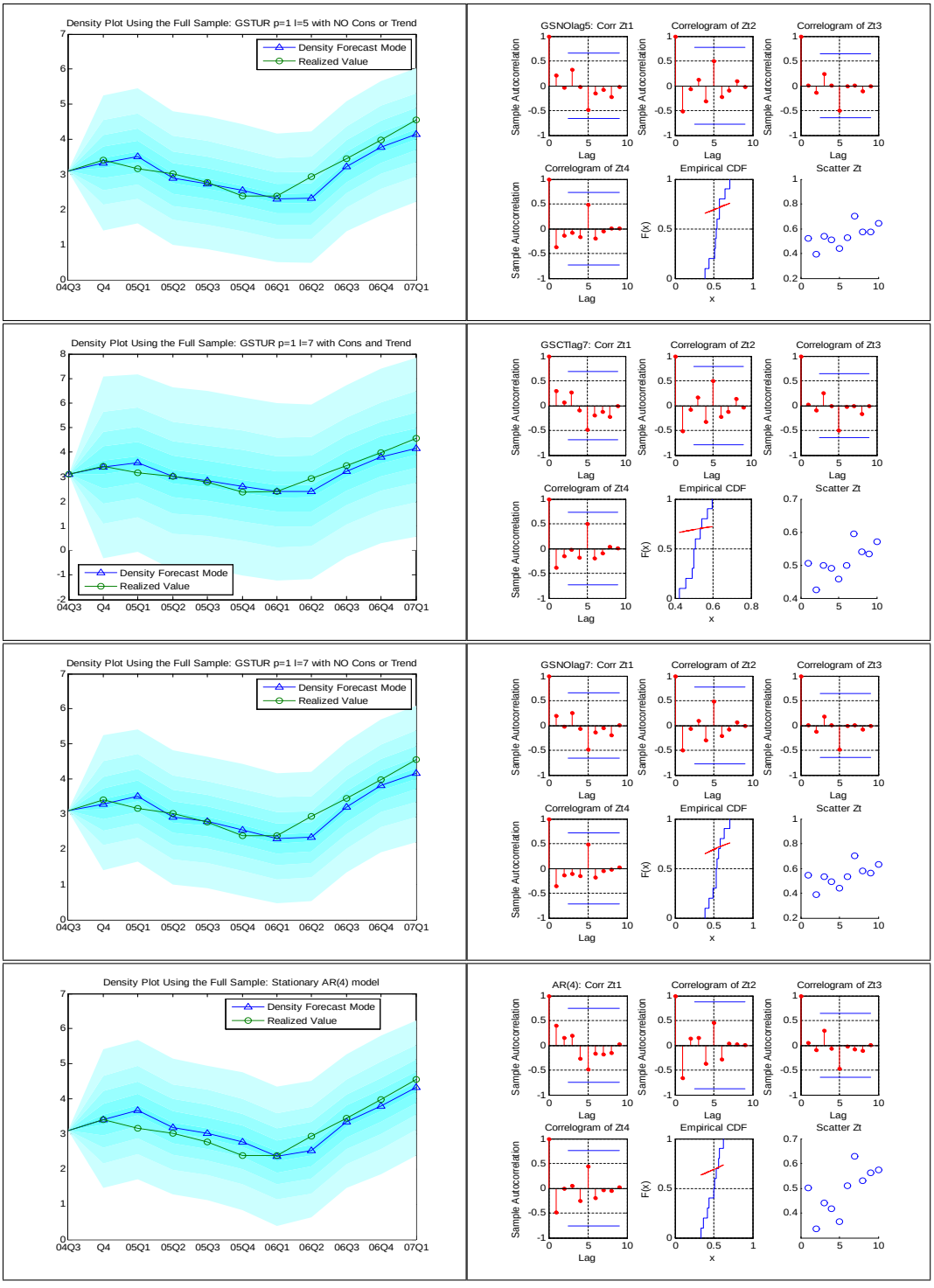
We provide the plots of the density forecast.


Density Plot Using the Full Sample: GSTUR p=1 l=0 with Cons
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSClag0: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: GSTUR p=1 l=0 with Trend
GSTlag0: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=0 with Cons and Trend
GSCTlag0: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=0 with NO Cons or Trend
GSNOlag0: Corr Zt1


Density Plot Using the Full Sample: GSTUR p=1 l=1 with Cons
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSClag1: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: GSTUR p=1 l=1 with Trend
GSTlag1: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=1 with Cons and Trend
GSCTlag1: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=1 with NO Cons or Trend
GSNOlag1: Corr Zt1


Density Plot Using the Full Sample: GSTUR p=1 l=4 with Cons
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSClag4: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: GSTUR p=1 l=4 with Trend
GSTlag4: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=4 with Cons and Trend
GSCTlag4: Corr Zt1
GSNOlag4: Corr Zt1


Density Plot Using the Full Sample: GSTUR lag4 No CT with APF
Density Forecast Mode
Realized Value
GS4NoAPF: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Density Plot Using the Full Sample: GSTUR p=1 l=5 with Cons
GSClag5: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=5 with Trend
GSTlag5: Corr Zt1
Density Plot Using the Full Sample: GSTUR p=1 l=5 with Cons and Trend
GSCTlag5: Corr Zt1
Sample Autocorrelation
Lag
F(x)
x
04Q3
Q4
05Q1
05Q2
05Q3
05Q4
06Q1
06Q2
06Q3
06Q4
07Q1


Density Plot Using the Full Sample: GSTUR p=1 l=5 with NO Cons or Trend
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSNOlag5: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: GSTUR p=1 l=7 with Cons and Trend
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSCTlag7: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: GSTUR p=1 l=7 with NO Cons or Trend
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GSNOlag7: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Full Sample: Stationary AR(4) model
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
AR(4): Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x

5.B.2 Small Sample

Forecast inflation Rates with Small Sample GSTUR p=1 lag=0 with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.0545	2.4312	2.8395	4.3050	-0.2534	7.4363	-32.2040
05Q1	3.17	3.4689	2.7373	3.2527	4.5577	0.1402	8.0896	-33.0112
05Q2	3.01	3.0800	2.7271	2.9054	3.3845	0.0733	7.0762	-34.3422
05Q3	2.77	2.8763	2.4383	2.7519	3.0888	-0.0355	6.5654	-35.1625
05Q4	2.38	2.5942	2.4274	2.5035	2.2265	-0.0466	5.8155	-35.9450
06Q1	2.39	2.2150	1.8087	2.1324	1.8645	-0.2352	5.0949	-37.0182
06Q2	2.93	2.2965	2.2823	2.2208	1.8417	-0.0877	5.1528	-37.4900
06Q3	3.44	2.9102	2.9691	2.8036	2.4660	0.2433	6.2281	-38.6328
06Q4	3.99	3.4773	3.0456	3.3555	2.6262	0.6921	7.0780	-39.5946
07Q1	4.55	4.0922	3.9788	3.9288	3.0096	1.1377	8.0517	-40.5594

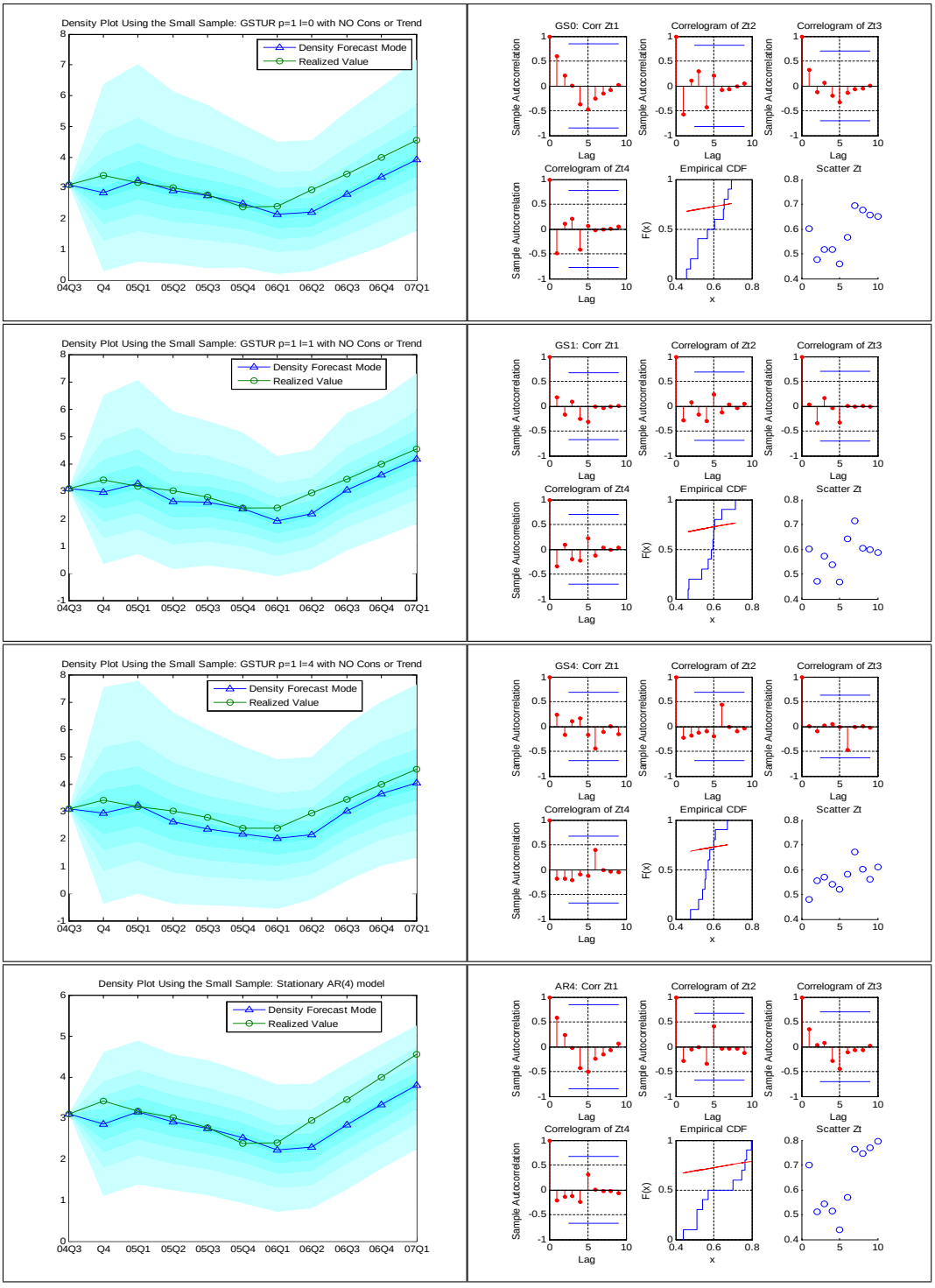
Forecast inflation Rates with Small Sample GSTUR p=1 lag=1 with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.1574	3.2501	2.9564	4.4825	-0.1808	7.7412	-38.2631
05Q1	3.17	3.5241	2.7048	3.2801	4.8258	0.1385	8.2304	-39.2315
05Q2	3.01	2.7856	2.7526	2.6089	3.7863	-0.2922	6.9902	-40.2030
05Q3	2.77	2.7391	2.1073	2.5979	2.9425	-0.1823	6.4681	-41.2633
05Q4	2.38	2.4579	2.3415	2.3482	2.7793	-0.3234	5.8461	-42.1774
06Q1	2.39	1.9733	1.8579	1.8997	1.9628	-0.5193	4.8897	-42.8431
06Q2	2.93	2.2401	2.0867	2.1827	1.9226	-0.2562	5.1257	-43.8280
06Q3	3.44	3.1443	2.9880	3.0372	2.5106	0.3641	6.5680	-45.0302
06Q4	3.99	3.6924	3.7626	3.5890	2.6398	0.8593	7.2489	-45.5372
07Q1	4.55	4.3245	3.9763	4.1569	3.0185	1.3652	8.1839	-46.6382

Forecast inflation Rates with Small Sample GSTUR p=1 lag=5 with NO CT								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	3.2561	5.1417	2.8468	21.3724	-1.5609	9.9720	-64.6217
05Q1	3.17	3.4125	2.0324	3.0389	10.8508	-1.1896	9.8626	-65.4927
05Q2	3.01	2.6349	0.4288	2.3428	18.3161	-1.9231	8.3540	-66.6004
05Q3	2.77	2.5214	4.8133	2.2770	11.6798	-1.5637	7.5988	-67.8078
05Q4	2.38	2.3147	1.3584	2.1516	5.7208	-1.4844	6.8658	-69.1216
06Q1	2.39	2.0780	3.1985	1.9806	4.6012	-1.3507	5.9576	-69.6748
06Q2	2.93	2.1083	1.9247	2.0030	3.2777	-1.1562	5.9027	-70.2495
06Q3	3.44	3.1216	3.0865	2.9927	4.0439	-0.4311	7.3652	-71.6289
06Q4	3.99	3.8087	3.3780	3.6593	4.4048	0.2230	8.2885	-72.4659
07Q1	4.55	4.1045	3.6833	3.8986	4.8095	0.5802	8.9850	-73.5496

Forecast inflation Rates with Small Sample AR(4)								
		Density forecast				95% percentile		
	TrueV	mean	mode	median	var	2.5%	97.5%	ML
04Q4	3.41	2.8463	2.8387	2.8489	1.1555	0.6969	4.9329	-25.6185
05Q1	3.17	3.1373	2.9423	3.1375	1.1492	0.9878	5.2432	-26.4645
05Q2	3.01	2.9034	2.7422	2.8998	1.0698	0.8638	4.9733	-27.1182
05Q3	2.77	2.7405	2.8613	2.7396	1.0026	0.7347	4.7427	-27.7514
05Q4	2.38	2.5256	2.5430	2.5195	0.9645	0.5869	4.4594	-28.3660
06Q1	2.39	2.2391	2.0545	2.2282	0.9026	0.3682	4.1263	-28.9833
06Q2	2.93	2.2923	2.3744	2.2961	0.8665	0.4766	4.1621	-29.5926
06Q3	3.44	2.8197	2.7332	2.8291	0.8792	0.9131	4.6301	-30.5750
06Q4	3.99	3.3100	3.2337	3.3318	0.8695	1.4330	5.1180	-31.4006
07Q1	4.55	3.7804	3.7858	3.7925	0.8622	1.8911	5.5541	-32.2413

Density Forecast Plots and PIT Evaluation

The following are the forecast density plots using small samples.


Density Plot Using the Small Sample: GSTUR p=1 l=0 with NO Cons or Trend
Density Forecast Mode
Realized Value
04Q3 Q4 05Q1 05Q2 05Q3 05Q4 06Q1 06Q2 06Q3 06Q4 07Q1
GS0: Corr Zt1
Correlogram of Zt2
Correlogram of Zt3
Correlogram of Zt4
Empirical CDF
Scatter Zt
Sample Autocorrelation
Lag
F(x)
x
Density Plot Using the Small Sample: GSTUR p=1 l=1 with NO Cons or Trend
GS1: Corr Zt1
Density Plot Using the Small Sample: GSTUR p=1 l=4 with NO Cons or Trend
GS4: Corr Zt1
Density Plot Using the Small Sample: Stationary AR(4) model
AR4: Corr Zt1

Chapter 6

Conclusions

This chapter firstly summarizes the work contained in this thesis. Our main findings are then reviewed and some suggestions for future research are made.

6.1 A Summary

This thesis relates to two bodies of literature. The first is a focus on inference in two nonlinear models, a generalization of Stochastic Unit Root (GSTUR) model and a Stationary Bilinear (SB) model, with a Bayesian approach. The second is a focus on forecasting in a Bayesian framework, in particular, density forecasting.

The main objective of this piece of research is to develop methods for Bayesian inference in the GSTUR and SB models, which includes model estimation, model comparison amongst non-nested models and constructing model-based forecasting. Procedures are developed, as by-products of this piece of research, to make the GSTUR and SB models readily applicable in a univariate time series modelling field. We developed three complete toolkits implemented in a Matlab environment. Software is prepared and ready to be disseminated for practitioners who wish to apply macroeconomic time series with a group of nonlinear models in the Bayesian framework. These toolkits include model estimation, model marginal likelihood calculation, forecasting and model comparison amongst GSTUR, SB, RW and stationary AR(p) models. All procedures and methodologies were intensively tested and validated on simulated data. Even with little knowledge of Bayesian statistics, following the instructions provided, economic practitioners could apply the GSTUR and SB models in a real setting.

One of the strengths of the toolkits we provide is that all the procedures are integrated in a Bayesian framework. Taking the toolkit of GSTUR for example, not only the high dimensional GSTUR model can be estimated, but also model marginal

likelihood can be evaluated for the purpose of model comparison, and a forecast density can be simulated with just a little extra cost on top of the estimation.

Compared with prior work in the study of GSTUR and SB models, the Markov chain Monte Carlo (MCMC) methods we proposed can produce efficient estimates with smaller variance. The Bayesian framework provides coherent, unified and operational approaches where no special techniques or principles are required for different problems. In particular, the Bayesian approach has more obvious advantages when the data is sparse (see chapter 2 for details). With the use of "model probabilities", comparing models amongst a pool of nonlinear and linear models is intuitively straight forward, whereas it is difficult to compare non-nested models in a classical framework.

Summarizations of each chapter are as follows:

In chapter 2, we provide the survey of Bayesian techniques at a high level of abstraction. We firstly explain the difference between Classical frequentist and Bayesian philosophy in understanding probabilities. Then, with respect to econometric modelling, we review the methodologies of model estimation, model comparison and model forecasting in both Classical and Bayesian framework. The logic of Bayesian econometrics reviewed in chapter 2 has provided a template for the organization of chapter 3, 4 and 5. To motivate the models we proposed, each chapter begins with an introduction of some puzzles in macroeconomic time series modelling. Then, likelihood functions, priors and derived posteriors are given. Computational methods for posterior inference and model comparison are then illustrated in details. Finally, we validate all methodologies on simulated data followed by applications in real settings.

In chapter 3, we extend a Stochastic Unit Root model to a Generalised STUR (GSTUR) model. According to the template provided in chapter 2, we overcome the estimation difficulty, and make applications to S&P 500 stock price data and U.K./U.S. real exchange rate data. We propose Chib method with an Auxiliary Particle Filter (APF) to obtain marginal likelihood when the entertained model is highly dimensional.

In chapter 4, we focus on a Stationary Bilinear model. Following the same procedure in chapter 3, we made Bayesian inference in the SB models and applied the developed toolbox to the quarterly UK inflation rates.

Finally, in chapter 5, we pool a group of nonlinear and linear models together and evaluate their forecasting capacities in a real setting. We also look into the field of combining forecasts, and compared the forecast results from a Bayesian Model Averaging (BMA) approach and a Simple Averaging (SA) approach with the results from individual forecasting models.

6.2 Main Findings and Suggestions for Future Research

This thesis sought answers in regard to both theoretical and empirical time series modelling. Apart from summarizing the main findings, I conclude this thesis with thoughts about directions for future research, and for promising applications.

In chapter 3, a GSTUR model is fitted to a macroeconomic time series and a financial time series in a Bayesian context. We applied the Bayesian framework because MCMC methods are good in dealing with high dimensional models. Bayes Factors in dealing with non-nested model comparison and Bayesian simulation approaches in dealing with density forecasts also show great promise. On reviewing the studies of Nelson and Plosser's data, we find that most of the debates have concerned whether the Nelson and Plosser's data series are trend-stationary or difference-stationary. With use of the Bayes Factors, we find strong evidence of a deterministic time trend when the underlying properties are allowed to change. Also in chapter 3, the Purchasing Power Parity theory is tested with an application of the UK/US long run real exchange rates. We suggest that the flexible GSTUR model, which allows the persistence to be changed at unknown times, may provide new insights in the PPP puzzle, and reconcile the empirical findings of non-stationarity in the literature with economic theories.

In the GSTUR model, the coefficient of the deterministic time trend is time-invariant. It is possible that this coefficient has varied at times due to the changes in technology and legislations. In this regard, Bayesian inference in a more Generalised STUR model is a promising area of research.

In chapter 4, by fitting a SB model and a RW model with the quarterly UK inflation data series, we find that a SB model better represents the underlying process of inflation according to the Bayes Factor.

Univariate inflation forecasting is the subject of much ongoing research. Forecasting inflation is difficult, because the inflation rate data is subject to a number of complexities including structural breaks, non-linear effects and measurement errors. For more flexibility, we may therefore extend the simplest SB model:

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t, \tag{6.1}$$

where $\varepsilon_t \sim i.i.d.f_N\left(0, \sigma_\varepsilon^2\right)$, to a two-threshold bilinear (TB) model:

$$y_t = (a_k + b_k\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t \tag{6.2}$$

where $\varepsilon_t \sim i.i.d. f_N\left(0, \sigma^2_{\varepsilon,k}\right)$ and $k = 1, 2$. This extended TB model incorporates breaks and allows for a change in volatilities, and hence, could be more adequate to serve the purpose of near-term inflation modelling compared with conventional linear forecasting models. Forecasting risks due to unexpected systematic shocks and/or disruptions in the underlying process of inflation caused by the breaks can then be inferenced from the extended TB model. Future research may focus on developing efficient algorithms to estimate a two-threshold bilinear model, and compare its forecasting capacity with a Stochastic Volatility model. Moreover, a great deal of work can be done in testing if the TB model encompasses other statistical forecasting models.

Chapter 4 also touches on the problem of sampling extreme values from the tails of a univariate normal distribution and a Gamma distribution by approximating the tail regions with an exponential distribution. We are inspired by the fact that in the real world, extreme events happen, such as the financial crisis, but how to model these extreme events is full of ambiguities. Thus, incorporating the extreme value theories in the Bayesian framework seems like a promising field of research for the following reasons: (1) The number of observations of the extreme events are relatively small. Therefore, we may ask whether the frequentist framework is fully justified when the realized data is sparse. (2)The Bayesian framework provides good solutions to missing data problems. However, there is a big question concerning how to construct our prior belief in the tail regions without encountering the criticism of being "subjective". Motivated by these two reasons, further research may focus on modelling the behaviour of the tail regions with applications to the financial time series data in a Bayesian framework.

In chapter 5, we focus on one-step ahead out-of-sample forecasts, including point forecast and density forecast. We find that GSTUR models with a constant provide better point forecasts with applications of the quarterly UK inflation rates (1957Q1-2007Q1), and also provide smaller Mean Squared Forecast Errors (MSFE). Therefore, we propose inflation modelers to take the GSTUR model into consideration. In more extensive analysis of statistical models' forecast capacities, we take the last 30 observations (1999Q4-2007Q1) as a subsample from the full data series. The GSTUR model outperforms the stationary AR(p) model with lag length selected with AIC. However, compared to forecasting models themselves, density forecasts using a small sample have much larger variances than those using the full sample. Our explanation is that simply abandoning part of the data may induce big forecast uncertainties, even though the dynamics of the underlying process might have changed. Therefore, in practice, we may start off using the full sample of data. Then, compare the forecast results from

those using a small sample. It is worth mentioning that Hashem Pesaran also raise the question as to the best sample period to use in estimation in his recent research (see Assenmacher- Wesche and Pesaran 2008, Pesaran and Timmermann 2007, Pesaran et al. 2006). His main solution is to estimate models with different observation windows and then pooling the forecasts. He proposes this method based on the idea that "estimation is more efficient if all available data are used when the models are stable" and "the occurrence of structural breaks might bias the forecasts" (Assenmacher-Wesche and Pesaran, 2008). Taking the breaks and model instability problem into account, he suggests to choose different sizes of observation windows to estimate, "starting from a minimum window size to the largest permitted by the available data set, and then average the forecasts across the windows" (Pesaran and Timmermann 2007). Since there are only two observation window sizes in our case (201 and 30), we may adopt Pesaran's approach to nonlinear forecasting models in future research and compare this "Averaging Window" forecasts with Bayesian Model Averaging approach to forecast combination.

Considering the issues regarding model selection and combining forecasts using Bayesian Model Averaging (BMA), only 24 models are considered in chapter 5, which includes GSTUR class models, a first order SB model, stationary AR(p) models and a RW model. Although the model comparison procedure using the marginal likelihood is universal from any kind of competing models, the calculation of each entertained model's marginal likelihood consumes a lot of time. The choice of the model dimension is then constrained by the computing resources. Clearly, a great deal of work is needed to develop efficient algorithms for comparing a large number of non-nested models. Also, as suggested in Armstrong (1978), we may combine the forecasts from econometric models with forecasts from other methods, such as forecasts from survey data, and see if the combination could yield improvements in accuracy.

Also in chapter 5, we investigate the forecasting capacities of a group of competing models and the main focuses are the GSTUR class models and a SB model. Since we could neither tell what exactly the underlying process is whereby we have those patterns of realizations, nor are we able to predict what exactly is going to happen in the future[1], the best we can do is to investigate some flexible models, which are more resilient to the changes and are able to accommodate the properties of the conventional linear and/or nonlinear time series models. Based on this motivation, further research can be performed by looking into various nonlinear models and their generalizations

[1]The two main sources that cause forecast failures: (1) things we do not know that we do not know, and (2) model uncertainties. Refer to chapter 5.1 for details.

with applications to different data sets. This motivation then will lead to research into evaluating the forecast capacities of the entertained nonlinear models under different circumstances, such as applications in financial modelling with high frequency data, and applications in macroeconomic modelling with low frequency data. One important application will be forecasting the UK inflation rate. With more sophisticated methods and flexible models, predictive densities can be compared to the fan charts published by the Bank of England. The results can then be compared with the findings in Cogley, Morozov and Sargent (2008).

In summary, nonlinear time series modelling is an under-developed field in applied Bayesian statistics. Not surprisingly, a wide variety of models and methods have not been fully investigated yet. In this sense, significant work in numerical analysis needs to be devoted to understanding the properties of algorithms and developing statistical software for econometric modelers. In general, future work contains generalising more flexible GSTUR and SB models and applying them to macroeconomic data, such as output gaps, unemployment rates and Business cycles.

Appendix A

Marginal Likelihood with Random Walk Model

The Random Walk (RW) model can be expressed as

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ \text{where } \varepsilon_t &\sim f_N\left(0, \sigma_\varepsilon^2\right) \end{aligned}$$

If we denote the error precision as h_ε, where $h_\varepsilon = \sigma_\varepsilon^{-2}$ and the prior chosen for $p\left(\sigma_\varepsilon^2\right)$ follows an Inverse-Gamma distribution, where $\sigma_\varepsilon^2 \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$. The Inverse-Gamma density can be expressed as:

$$p\left(\sigma_\varepsilon^2\right) = \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma\left(\underline{\alpha}_\varepsilon\right)\left(\sigma_\varepsilon^2\right)^{\underline{\alpha}_\varepsilon+1}} \exp\left\{-\frac{1}{\underline{\beta}_\varepsilon \sigma_\varepsilon^2}\right\}$$

If we denote $\overline{\beta}_\varepsilon^{-1} = \frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2}\sum_{t=2}^{N}\left(y_t - y_{t-1}\right)^2$, the marginal likelihood $p\left(y \mid M_{RW}\right)$ then

can be calculated as the following:

$$\begin{aligned}
p(y \quad | \quad M_{RW}) &= \int p(y \mid \sigma_\varepsilon^2) p(\sigma_\varepsilon^2) d\sigma_\varepsilon^2 \\
&= \int \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{\frac{N-1}{2}}} \exp\left\{ -\frac{\sum_{t=2}^{N} \left(y_t - y_{t-1}\right)^2}{2\sigma_\varepsilon^2} \right\} \\
&\quad \cdot \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right) \left(\sigma_\varepsilon^2\right)^{\underline{\alpha}_\varepsilon + 1}} \exp\left\{ -\frac{1}{\underline{\beta}_\varepsilon \sigma_\varepsilon^2} \right\} d\sigma_\varepsilon^2 \\
&= \int \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right) (2\pi)^{\frac{N-1}{2}} \left(\sigma_\varepsilon^2\right)^{\frac{N+2\underline{\alpha}_\varepsilon+1}{2}}} \exp\left\{ -\frac{1}{\overline{\beta}_\varepsilon \sigma_\varepsilon^2} \right\} d\sigma_\varepsilon^2 \\
&= \int \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right) (2\pi)^{\frac{N-1}{2}}} h_\varepsilon^{\frac{N+2\underline{\alpha}_\varepsilon+1}{2}} \exp\left\{ -\frac{h_\varepsilon}{\overline{\beta}_\varepsilon} \right\} d\left(\frac{1}{h_\varepsilon}\right) \\
&= -\frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right) (2\pi)^{\frac{N-1}{2}}} \int h_\varepsilon^{\frac{N+2\underline{\alpha}_\varepsilon-3}{2}} \exp\left\{ -\frac{h_\varepsilon}{\overline{\beta}_\varepsilon} \right\} dh_\varepsilon \\
&= \frac{\overline{\beta}_\varepsilon^{\overline{\alpha}_\varepsilon} \Gamma\left(\overline{\alpha}_\varepsilon\right)}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right) (2\pi)^{\frac{N-1}{2}}}
\end{aligned}$$

where

$$\overline{\alpha}_\varepsilon = \underline{\alpha}_\varepsilon + \frac{N-1}{2}$$

and

$$\overline{\beta}_\varepsilon = \left[\frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2} \sum_{t=2}^{N} \left(y_t - y_{t-1}\right)^2 \right]^{-1}$$

Integrating out nuisance parameters has to be executed with caution as emphasized by Box and Tiao (1992, pp.71). The values of $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ in $p\left(\sigma_\varepsilon^2\right) \sim f_{\Gamma_\varepsilon}^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ determines the marginal likelihood value $p(y)$. Thus, $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ have to be chosen as appropriate values to reflect the data information sufficiently. Choosing $\underline{\alpha}_\varepsilon$ and $\underline{\beta}_\varepsilon$ a flat prior arbitrarily may be inappropriate and, as a result, provide misleading inference in the model selection stage.

Appendix B

Probability Distributions

To clarify calculation issues involved with conjugate priors, we provide a summarization of some probability density functions. Zellner (1971, Appendix A) provides details of the properties of various probability distributions, including mean, variance, skewness and kurtosis. Bauwens et al. (1999) Appendix A provide good summaries in this regard as well. More over, they also provide algorithms of how to generate random numbers from a probability distribution in Appendix B. In this appendix, we follow the Appendix 1 in Berger (1985), regroup the definitions of the distributions that are intensively applied in this thesis.

Gamma Distribution: $x \sim f_{\Gamma}(\alpha, \beta)$

$$p(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$$

$mean = \alpha\beta$

$Variance = \alpha\beta^2$

Inverse Gamma Distribution: $x \sim f_{\Gamma}^{-1}(\alpha, \beta)$

$$p(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha) x^{\alpha+1}} \exp\left(-\frac{1}{x\beta}\right)$$

$mean = 1/\beta(\alpha - 1)$ if $\alpha > 1$

$Variance = 1/\beta^2(\alpha - 1)^2(\alpha - 1)$ if $\alpha > 2$

Note: $1/X \sim f_{\Gamma}(\alpha, \beta)$

Multivariate t distribution:

p-variate t distribution with α degree of freedom, location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$: $p \times p$ positive definite matrix, and

$$p(x) = \frac{\Gamma[(\alpha+p)/2]}{(\det \mathbf{\Sigma})^{1/2}(\alpha\pi)^{p/2}\Gamma(\alpha/2)}\left[1+\frac{1}{\alpha}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-(\alpha+p)/2}$$

$mean = \boldsymbol{\mu}$ if $\alpha > 1$

$Covariance = \alpha\mathbf{\Sigma}/(\alpha-2)$ if $\alpha > 2$

Multivariate normal distribution:

matrix $\mathbf{\Sigma}$: $p \times p$ positive definite matrix, and

$$p(x) = \frac{1}{(\det \mathbf{\Sigma})^{1/2}(2\pi)^{p/2}}\exp\left\{-.5(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

$mean = \boldsymbol{\mu}$

$Covariance = \mathbf{\Sigma}$

BIBLIOGRAPHY

Bibliography

[1] Armstrong, J.S. (1978). Econometric Forecasting and the Science Court. Journal of Business, 51(4), 595-600.

[2] Assenmacher-Wesche, K. and Pesaran, M.H. (2008). Forecasting the Swiss Economy Using VECX Models: An Exercise in Forecast Combination Across Modelsand Observation Windows. Working Papers 2008-3, Swiss National Bank.

[3] Atkeson, A. and Ohanian, L.E. (2001). Are Phillips Curves Useful for Forecasting Inflation? FRB Minneapolis Quarterly Review, 25(1), 2-11.

[4] Baillie, R.T., Chung, C. and Tieslau, M.A. (1996). Analysing Inflation by the Fractionally Integrated Arfima–Garch Model. Journal of Applied Econometrics, 11(1), 23-40.

[5] Bartlett, M.S. (1957). A Comment on D. V. Lindley's Statistical Paradox. Biometrika, 44(3/4), 533-4.

[6] Bates, J.M. and Granger, C.W.J. (1969). The Combination of Forecasts. Operational Research Quarterly, 20, 451-68.

[7] Baum, C., Barkoulas, J. and Caglayan, M. (2001). Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. Journal of International Money and Finance, 20(3), 379-99.

[8] Bauwens, L. (1999). Bayesian inference in dynamic econometric models. Oxford: Oxford University Press.

[9] Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. Econometrics Journal, 1(ConferenceIssue), C23-46.

[10] Berger, J.O. (1985). Statistical decision theory and Bayesian analysis. 2nd ed. New York: Springer-Verlag.

[11] Bernanke, B.S. (2007). Inflation Expectations and Inflation Forecasting. At the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts.

[12] Bernardo, J.M. and Smith, A.F.M. (1994). Bayesian theory. Chichester: Wiley.

[13] Bleaney, M.F., Leybourne, S.J. and Mizen, P. (1999). Mean reversion of real exchange rates in high-inflation countries. Southern Economic Journal.

[14] Bos, C.S. (2002). A Comparison of Marginal Likelihood Computation Methods. Tinbergen Institute Working Paper no. 2002-084/4.

[15] Box, G.E.P. and Tiao, G.C. (1992). Bayesian inference in statistical analysis. Wiley.

[16] Britton, E., Fisher, P. and Whitley, J. (February 1998). The Inflation Report Projections: Understanding the Fan Chart. Bank of England Quarterly Bulletin, 30-7.

[17] Broock, W.A., Scheinkman, J. A., Dechert, W. D. and LeBaron, B. (1996). A test for independence based on the correlation dimension. Econometric Reviews, 15(3), 197.

[18] Brunner, A.D. and Hess, G.D. (1995). Potential problems in estimating bilinear time-series models. Journal of Economic Dynamics and Control, 19(4), 663-81.

[19] Bunn, D.W. (1985). Statistical efficiency in the linear combination of forecasts. International Journal of Forecasting, 1(2), 151-63.

[20] Busetti, F. and Taylor, A.M.R. (2004/11). Tests of stationarity against a change in persistence. Journal of Econometrics, 123(1), 33-66.

[21] Byers, D. and Peel, D.A. (1995). Bilinear Quadratic ARCH and Volatility Spillovers in Inter-war Exchange Rates. Applied Economics Letters., 2(7), 215-9.

[22] Caner, M. and Hansen, B.E. (2001). Threshold Autoregression with a Unit Root. Econometrica, 69(6), 1555-96.

[23] Canova, F. (1993). Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model. Journal of Economic Dynamics and Control, 17(1-2), 233-61.

[24] Carlin, B.P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society. Series B (Methodological), 57(3), 473-84.

[25] Carlin, B.P. and Louis, T.A. (2000). Bayes and empirical bayes methods for data analysis. 2nd ed. Boca Raton; London: Chapman & Hall/CRC.

[26] Carpenter, J., Clifford, P. and Fearnhead, P. (1999). An improved particle filter for non-linear problems. IEE Proceedings - Radar, Sonar and Navigation, 146, 2-7.

[27] Caves, R.E. (1980). Britain's economic performance. Washington: DC Brookings Institution.

[28] Charemza, W.W., Kharin, Y. and Maevskiy, V. (2006). Bilinear forecast risk assessment for non-systematic inflation: Theory and evidence. National Bank of Poland Working Paper.

[29] Charemza, W.W., Lifshits, M. and Makarova, S. (2005). Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results. Journal of Economic Dynamics and Control, 29(1-2), 63-96.

[30] Chatfield, C. (1993). Calculating Interval Forecasts. Journal of Business & Economic Statistics, 11(2), 121-35.

[31] Chatfield, C. (1988). The future of the time-series forecasting. International Journal of Forecasting, 4(3), 411-9.

[32] Chib, S. (1995). Marginal Likelihood from the Gibbs Output. Journal of the American Statistical Association, 90(432), 1313-21.

[33] Chib, S. (1993). Bayes regression with autoregressive errors : A Gibbs sampling approach. Journal of Econometrics, 58(3), 275-94.

[34] Chib, S. and Jeliazkov, I. (2005). Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. Statistica Neerlandica, 59(1), 30-44.

[35] Chib, S., Nardari, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. Journal of Econometrics, 134(2), 341-71.

[36] Chib, S., Nardari, F. and Shephard, N. (2002/6). Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics, 108(2), 281-316.

[37] Christoffersen, P.F. (1998). Evaluating Interval Forecasts. International Economic Review, 39(4), 841-62.

[38] Christoffersen, P.F. and Diebold, F.X. (1997). Optimal Prediction Under Asymmetric Loss. Econometric Theory, (13), 808-17.

[39] Clemen, R.T. (1986). Linear constraints and the efficiency of combined forecasts. Journal of Forecasting, 5, 31-8.

[40] Clemen, R.T. and Winkler, R.L. (Jan., 1986). Combining Economic Forecasts. Journal of Business & Economic Statistics, 4(1), 39-46.

[41] Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4), 559-83.

[42] Clements, M.P. (2006). Evaluating the survey of professional forecasters probability distributions of expected inflation based on derived event probability forecasts. Empirical Economics, 31(1), 49-64.

[43] Clements, M.P. (2004). Evaluating the Bank of England Density Forecasts of Inflation. The Economic Journal, 114(498), 844-66.

[44] Clements, M.P., Galvao, A.B. and Kim, J.H. (2006). Quantile Forecasts of Daily Exchange Rate Returns from Forecasts of Realized Volatility. The Warwick Economics Research Paper Series (TWERPS) 777, University of Warwick, Department of Economics.

[45] Clements, M.P. and Hendry, D.F. (1998). Forecasting economic time series. Cambridge: Cambridge University Press.

[46] Clements, M.P. and Smith, J. (2000). Evaluating the forecast densities of linear and non linear models applications to output growth and unemployment. Journal of Forecasting, (19), 255-76.

[47] Clyde, M., Desimone, H. and Parmigiani, G. (1996). Prediction Via Orthogonalized Model Mixing. Journal of the American Statistical Association, 91(435), 1197-208.

[48] Cogley, T. (2005). Inflation Dynamics. University of California, Davis, Working Paper.

[49] Cogley, T. and Sargent, T. (2001). Evolving Post World War II US Inflation Dynamics. NBER Macroeconomics Annual, 331-72.

[50] Cogley, T. and Sargent, T. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War US. Review of Economic Dynamics, 8, 262-302.

[51] Cogley, T., Primiceri, G.E. and Sargent, T.J. (2008). Inflation-Gap Persistence in the U.S. NBER Working Papers 13749, National Bureau of Economic Research.

[52] Cowles, M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. Journal of the American Statistical Association, 91(434), 883-904.

[53] Davis, R.A. and Resnick, S.I. (1996). Limit Theory for Bilinear Processes with Heavy-Tailed Noise. The Annals of Applied Probability, 6(4), 1191-210.

[54] Dawid, A.P., Stone, M. and Zidek, J.V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). Journal of Royal Statistics Society Series, 35, 189-233.

[55] de Menezes, L.M., W. Bunn, D. and Taylor, J.W. (2000). Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120(1), 190-204.

[56] DeJong, D.N. and Whiteman, C.H. (1991 a). The Case for Trend-Stationarity is Stronger than we Thought. Journal of Applied Econometrics, 6(4), 413-21.

[57] DeJong, D.N. and Whiteman, C.H. (1991 b). Reconsidering 'trends and random walks in macroeconomic time series'. Journal of Monetary Economics, 28(2), 221-54.

[58] Deutsch, M., Granger, C.W.J. and Teravirta, T. (1994). The combination of forecasts using changing weights. International Journal of Forecasting, 10(1), 47-57.

[59] Devroye, L. (1986). Non-uniform random variate generation. New York: Springer.

[60] Diebold, F.X., Gunther, T.A. and Tay, A.S. (May 1997). Evaluating density forecasts. National Bureau of Economic Research, Working Papers 97-6.

[61] Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39(4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance), 863-83.

[62] Diebold, F.X., Hahn, J. and Tay, A.S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange. The Review of Economics and Statistics, 81(4), 661-73.

[63] Diebold, F.X. and Lopez, J.A. (1996). Forecast Evaluation and Combination. In Maddala, G.S. andRao, C.R. (eds.), Handbook of Statistics 14: Statistical Methods in Finance, Amsterdam: North-Holland.: 241-68.

[64] Dowd, K. (2007). Too good to be true? The (In)credibility of the UK inflation fan charts. Journal of Macroeconomics, 29(1), 91-102.

[65] Draper, D. (1999). Discussion of "Decision models in screening for breast cancer" by G. Parmigiani. Bayesian Statistics 6, 541-3.

[66] Eklund, J. and Karlsson, S. (2005). Forecast Combination and Model Averaging using Predictive Measures. Working Paper Series 191, Sveriges Riksbank (Central Bank of Sweden).

[67] Elliott, G., Granger, C.W.J. and Timmermann, A. (2006). Handbook of economic forecasting. Amsterdam; London: Elsevier North Holland.

[68] Enders, W. (1995). Applied econometric time series. New York; Chichester: Wiley.

[69] Engel, C. and Kim, C. (1999). The Long-Run U.S./U.K. Real Exchange Rate. Journal of Money, Credit and Banking, 31(3, Part 1), 335-56.

[70] Faria, A.E. and Souza, R.C. (1995). A re-evaluation of the quasi-bayes approach to the linear combination of forecasts. Journal of Forecasting, 14(6), 533-42.

[71] Fernandez, C., Ley, E. and Steel, M.F.J. (2001). Model Uncertainty in Cross-Country Growth Regressions. Journal of Applied Econometrics, 16(5), 563-76.

[72] Fearnhead P. (2004). Particle filters for mixture models with an unknown number of components. Statistics and Computing, 14, 11-21.

[73] Fearnhead P. (2002). MCMC, sufficient statistics and particle filter. Journal of Computational and Graphical Statistics, 11, 848-62.

[74] Fisher, J.D.M., Liu, C.T. and Zhou, R. (2002). When can we forecast inflation? Economic Perspectives, (1), 32-44.

[75] Fox, R. and Taqqu, M.S. (1986). Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series. The Annals of Statistics, 14(2), 517-32.

[76] Francq, C., Makarova, S. and Zakoın, J. (2008). A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test. Journal of Econometrics, 142(1), 312-26.

[77] Frankel, J.A. (1994). Exchange Rate Policy In American Economic Policy in the 1980s. In Martin Feldstein (ed.), Chicago: NBER and University of Chicago Press.

[78] Frankel, J. and Rose, A. (1996). A panel project on purchasing power parity: Mean reversion within and between countries. Journal of International Economics, 40(1-2), 209-24.

[79] Friedman, M. and Schwartz, Anna J. A Monetary History of the United States, 1867-1960. In Princeton University Press.

[80] Garratt, A., Mise, Emi, Koop, Gary and Vahey, Shaun P. (2007). Real-time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty. Birkbeck College Working Paper.

[81] Geisser, S. (1984). On Prior Distributions for Binary Trials. The American Statistician, 38(4), 244-7.

[82] Gelfand, A.E. and Dey, D.K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. Journal of the Royal Statistical Society.Series B (Methodological), 56(3), 501-14.

[83] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association, 85(410), 398-409.

[84] Gelman, A. (1995). Bayesian data analysis. 1st ed. London; New York: Chapman & Hall.

[85] Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7(4), 457-72.

[86] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[87] Geweke, J. (1999). Using simulation methods for bayesian econometric models: inference, development,and communication. 18(1), 1-76.

[88] Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments Bayesian Statistics 4, 169-93.

[89] Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. In Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Alexandria, VA: American Statistical Association., 571–578.

[90] Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. Econometrica, 57(6), 1317-39.

[91] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. 4, 221-238.

[92] Gil-Ala, L. A.and Robinson, P. M. (1997). Testing of unit root and other non-stationary hypotheses in macroeconomic time series. Journal of Econometrics, 80(2), 241-268.

[93] Giraitis, L. and Surgailis, D. (2002). ARCH-type bilinear models with double long memory. Stochastic Processes and their Applications, 100(1-2), 275-300.

[94] Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Radar and Signal Processing, IEE Proceedings F, 140(2), 107-13.

[95] Granger, C.W.J. (2003, Dec 8). Time series analysis, cointegration, and applications. Nobel Lecture.

[96] Granger, C.W.J. and Anderson, A.P. (1978). An Introduction to Bilinear Time Series Models. Gottingen: Vandenhoeck and Ruprecht.

[97] Granger, C.W.J. and Ramanathan, R. (1984). Improved Methods of Combining Forecasts. Journal of Forecasting 3, 3, 197-204.

[98] Granger, C.W.J. and Swanson, N.R. (1997/9). An introduction to stochastic unit-root processes. Journal of Econometrics, 80(1), 35-62.

[99] Green, P.J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika, 82(4), 711-32.

[100] Grilli, V. and Kaminsky, G. (1991). Nominal Exchange Rate Regimes and the Real Exchange Rate: Evidence from the United States and Great Britain, 1885-1986.. Journal of Monetary Economics, 27, 191-212.

[101] Guerard, J.B. (1987). Linear Constraints, Robust-Weighting and Efficient Composite. Modeling. Journal of Forecasting, 6(3), 193-6.

[102] Gupta, S. and Wilton, P.C. (1987). Combination of Forecasts: An Extension. Management Science, 33(3), 356-72.

[103] Guyton, D.A., Zhang, N.. and Foutz, R.V. (1986). A random parameter process for modeling and forecasting time series. 7, 105–115.

[104] Hall, S.G. and Mitchell, J. (2007). Combining density forecasts. International Journal of Forecasting, 23(1), 1-13.

[105] Hall, S.G. and Mitchell, J. (2004). Density Forecast Combination. NIESR Discussion Paper no.249.

[106] Han, C. and Carlin, B.P. (2001). Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review. Journal of the American Statistical Association, 96(455), 1122-32.

[107] Hannan, E.J. (1982). A note on bilinear time series models. Stochastic Processes and their Applications, 12(2), 221-4.

[108] Hansen, B.E. (2006). Interval forecasts and parameter uncertainty. Journal of Econometrics, 135(1-2), 377-98.

[109] Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2006/10). Modified tests for a change in persistence. Journal of Econometrics, 134(2), 441-69.

[110] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), 97-109.

[111] Hendry, D.F. (2001). Modelling UK Inflation, 1875-1991. Journal of Applied Econometrics, 16(3, Special Issue in Memory of John Denis Sargan, 1924-1996: Studies in Empirical Macroeconometrics), 255-75.

[112] Hendry, D.F. and Clements, M.P. (2003). Economic forecasting: some lessons from recent research. Economic Modelling, 20, Issue 2, March 2003, Pages 301-329(2), 301-29.

[113] Hendry, D.F. and Ericsson, N.R. (2001). Understanding economic forecasts. Cambridge, Mass.: MIT Press.

[114] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. Statistical Science, 14(4), 382-401.

[115] Holden, K. and Peel, D.A. (1986). An empirical investigation of combinations of economic forecasts. Journal of Forecasting, 5, 229-42.

[116] Holmen, J.S. (1987). A note on the value of combining short-term earnings forecasts : A test of Granger and Ramanathan. International Journal of Forecasting, 3(2), 239-43.

[117] Hristova, D. (2005). Maximum Likelihood Estimation of a Unit Root Bilinear Model with an Application to Prices. Studies in Nonlinear Dynamics & Econometrics, 9(1), Pages 1199.

[118] Igloi, E. and Terdik, G. (1999). Bilinear Stochastic Systems with Fractional Brownian Motion Input. The Annals of Applied Probability, 9(1), 46-77.

[119] Jones, C.R. and Marriott, M.J. (1999). A Bayesian Analysis of Stochastic Unit Root Models. Bayesian Statistics, 6, 785-794.

[120] Judge, G.G., Hill, C.R., Griffiths, W.E., Lutkepohl, H., and Lee, T.C. (1985). The theory and practice of econometrics. 2nd ed. New York ; Chichester: Wiley.

[121] Kapetanios, G. (2004). A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset. Economics Letters, 85(1), 63-9.

[122] Kapetanios, G., Shin, Y. and Snell, A. (2003/2). Testing for a unit root in the nonlinear STAR framework. Journal of Econometrics, 112(2), 359-79.

[123] Karlsson, S. and Jacobson, T. (2004). Finding good predictors for inflation: a Bayesian model averaging approach. Journal of Forecasting, 23(7), 479-96.

[124] Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. Journal of the American Statistical Association, 90(430), 773-95.

[125] Kim, W.R., Billard, L. and Basawa, I.V. (1990). Estimation for the first-order diagonal bilinear time series model. Journal of Time Series Analysis, 11, 215-29.

[126] Kim, I. and Maddala, G.S. (1991). Flat Priors vs. Ignorance Priors in the Analysis of the AR(1) Model. Journal of Applied Econometrics, 6(4), 375-80.

[127] Kim, J. (2000/3). Detection of change in persistence of a linear time series. Journal of Econometrics, 95(1), 97-116.

[128] Kim, J. and Amador, R.B. (2002/8). Corrigendum to "Detection of change in persistence of a linear time series" [J. Econom. 95 (2000) 97-116]. Journal of Econometrics, 109(2), 389-92.

[129] Kim, Y. (1990). Purchasing Power Parity in the Long Run: A Cointegration Approach. Journal of Money, Credit and Banking, 22(4), 491-503.

[130] King, M. (1997). Changes in UK monetary policy: Rules and discretion in practice. Journal of Monetary Economics, 39(1), 81-97.

[131] Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. Journal of Computational and Graphical Statistics, 5(1), 1-25.

[132] Kontonikas, A. (2002). Inflation and Inflation Uncertainty in the United Kingdom: Evidence from GARCH modelling. Brunel University, School of Social Sciences, Working Paper.

[133] Koop, G., Poirier,Dale J. and Tobias,Justin L. (2007). Bayesian econometric methods. Cambridge: Cambridge University Press.

[134] Koop, G. (1994/7). An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates. Journal of Empirical Finance, 1(3-4), 343-64.

[135] Koop, G., Ley, Eduardo, O., J. and Steel, M.F.J. (1997/0). Bayesian analysis of long memory and persistence using ARFIMA models. Journal of Econometrics, 76(1-2), 149-69.

[136] Koop, G. and Potter, S.M. (2004 Dec). Forecasting and Estimating Multiple Change-Point Models with an UnknownNumber of Change Points. Federal Reserve Bank of New York Staff Reports, (no.196).

[137] Koop, G. and Potter, S.M. (1999). Bayes factors and nonlinearity: Evidence from economic time series. Journal of Econometrics, 88(2), 251-81.

[138] Koop, G. and Steel, M.F.J. (1991). A Comment on: 'To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends', by Peter C. B. Phillips. Journal of Applied Econometrics, 6(4), 365-70.

[139] Kwiatkowski, D., Phillips, Peter C.B., Schmidt, P., Shin,Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1-3), 159-78.

[140] Lane, J.A., Peel, D.A. and Raeburn, E.J. (1996). Some Empirical Evidence on the Time-Series Propertieicess of Four UK Asset Price. Economica, 63(251), 405-26.

[141] LeSage, J. (1998). Applied Econometrics using MATLAB. University of Toledo.

[142] LeSage, J.P. (1999). Econometrics Toolbox–Created using MATLAB. Spatial-Econometrics.

[143] Leybourne, S.J., McCabe, B. P. M. and Tremayne, A.R. (1996). Can Economic Time Series Be Differenced to Stationarity? Journal of Business and Economic Statistics, 14(4), 435-46.

[144] Li, K., Wong, H. and Troutt, M. (2001). An Approximate Bayesian Algorithm for Combining Forecasts. Decision Sciences, 32(3), 453-72.

[145] Li, W.K. and McLeod, A.I. (1986). Fractional Time Series Modelling. Biometrika, 73(1), 217-21.

[146] Liu, J.S. (2001). Monte Carlo strategies in scientific computing. New York ; London: Springer.

[147] Lobo, G.J. (1991). Alternative methods of combining security analysts' and statistical forecasts of annual corporate earnings. International Journal of Forecasting, 7(1), 57-63.

[148] Lothian, J.R. and Taylor, M.P. (1996). Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries. The Journal of Political Economy, 104(3), 488-509.

[149] Lothian, J. and Taylor, M. (1997). Real exchange rate behavior. Journal of International Money and Finance, 16(6), 945-54.

[150] Lucas, R.E.,Jr. (2000). Inflation and Welfare. Econometrica, 68(2), 247-74.

[151] Madigan, D. and York, J. (1995). Bayesian Graphical Models for Discrete Data. International Statistical Review / Revue Internationale De Statistique, 63(2), 215-32.

[152] Madigan, D. and Raftery, A.E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. Journal of the American Statistical Association, 89(428), 1535-46.

[153] Maravall, A. (1983). An Application of Nonlinear Time Series Forecasting. Journal of Business and Economic Statistics, 1, 66-74.

[154] Martins, C.M. (1997). A note on the autocorrelations related to a bilinear model with non-independent shocks. Statistics & Probability Letters, 36(3), 245-50.

[155] Meddeb, S., Tourneret, J.Y. and Castanie, F. (1998, May). Identification of bilinear systems using Bayesian inference. Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference, 3, 1609-12.

[156] Meese, R. and Geweke, J. (1984). A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. Journal of Business and Economic Statistics, 2(3), 191-200.

[157] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics, 21, 1087-92.

[158] Michael, P., Nobay, A.R. and Peel, D.A. (1997). Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. The Journal of Political Economy, 105(4), 862-79.

[159] Milani, F. (2004). Monetary Policy with a Wider Information Set: a Bayesian Model Averaging Approach. Macroeconomics 0401004, EconWPA.

[160] Milton Friedman (1977). Nobel Lecture: Inflation and Unemployment. The Journal of Political Economy, 85(3), 451-72.

[161] Mitchell, J. (2005). The National Institute density forecasts of inflation. National Institute Economic Review, 193, 60-9.

[162] Morris, C.N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. Journal of the American Statistical Association, 78(381), 47-55.

[163] Morris, P.A. (1977). Combining Expert Judgments: A Bayesian Approach. Management Science, 23(7), 679-93.

[164] Nelson, C.R. and Plosser, C.I., (1982). Trends and random walks in macroeconomic time series Journal of Monetary Economics, 10, 139-162.

[165] Newbold, P. and Granger, C.W.J. (1974). Experience with Forecasting Univariate Time Series and the Combination of Forecasts. Journal of the Royal Statistical Society.Series A (General), 137(2), 131-65.

[166] Newbold, P., Leybourne, S. and Wohar, M.E. (2001/0). Trend-stationarity, difference-stationarity, or neither: further diagnostic tests with an application to U.S. Real GNP, 1875-1993. Journal of Economics and Business, 53(1), 85-102.

[167] Newey, W.K. and West, K.D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703-8.

[168] Newton, M.A. and Raftery, A.E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. Journal of the Royal Statistical Society.Series B (Methodological), 56(1), 3-48.

[169] Obstfeld, M. and Taylor, A. (1997). Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited, Journal of the Japanese and International Economies, 11(4), 441-79.

[170] O'Connell, P.J. (1998). The overvaluation of purchasing power parity. Journal of International Economics, 44(1), 1-19.

[171] Osborn, D. and Sensier, M. (2007). UK inflation: persistance, seasonality and monetary policy. The School of Economics Discussion Paper Series 0716, Economics, the University of Manchester.

[172] Palm, F.C. and Zellner, A. (1992). To combine or not to combine: issues of combining forecasts. Journal of Forecasting, 11, 687-701.

[173] Papell, D. (1997). Searching for stationarity: Purchasing power parity under the current float. Journal of International Economics, 43(3-4), 313-32.

[174] Peel, D.A. and Davidson, J. (1998). A non-linear error correction mechanism based on the bilinear model. Economics Letters, 58(2), 165-70.

[175] Pesaran, M.H., Pettenuzzo, D. and Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. Review of Economic Studies, 73(4), 1057-84.

[176] Pesaran, M.H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics, 137(1), 134-61.

[177] Phillips, P.C.B. (1991). To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends. Journal of Applied Econometrics, 6(4), 333-64.

[178] Pitt, M.K. and Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association, 94(446), 590-9.

[179] Poirier, D.J. (1995). Intermediate statistics and econometrics : a comparative approach. Cambridge, Mass. ; London: MIT Press.

[180] Potter, S.M. (1999). Nonlinear Time Series Modelling: An Introduction. Journal of Economic Surveys, 13(5), 505-28.

[181] Potter, S. and Koop, G. (2003). Forecasting in large macroeconomic panels using Bayesian Model Averaging. Federal Reserve Bank of New York in its Series Staff Reports with Number 163.

[182] Potter, S., Koop, G. and Strachan, R.W. (2005). Re-examining the Consumption-Wealth Relationship: The Role of Model Uncertainty. FRB of New York Staff Report no. 202.

[183] Poulizac, D., Weale, M. and Young, G. (1996). The Performance of National Institute Economic Forecasts. National Institute Economic Review, 156(1), 55-62.

[184] Quinn, B.G. (1982). Stationarity and invertibility of simple bilinear models. Stochastic Processes and their Applications, 12(2), 225-30.

[185] Raftery, A.E., Balabdaoui, F. Gneiting, T. and Polakowski, M. (2003). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Technical Report no. 440, Department of Statistics, University of Washington.

[186] Rao, T.S. (1981). On the Theory of Bilinear Time Series Models. Journal of the Royal Statistical Society.Series B (Methodological), 43(2), 244-55.

[187] Rao, T.S. and Sesay, S. A. O. (1988). Yule–Walker type difference equations for higher-order moments and cumulants for the bilinear time series models. Journal of Time Series Analysis, 9, 385-401.

[188] Rapach, D.E. and Wohar, M.E. (2006). The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. International Journal of Forecasting, 22(2), 341-61.

[189] 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(419), 861-8.

[190] Rogoff, K. (1996). The Purchasing Power Parity Puzzle. Journal of Economic Literature, 34(2), 647-68.

[191] Sarantis, N. (1999). Modeling non-linearities in real effective exchange rates. Journal of International Money and Finance, 18(1), 27-45.

[192] Sarno, L. (2005). Viewpoint: Towards a solution to the puzzles in exchange rate economics: where do we stand? Canadian Journal of Economics/Revue Canadienne d'Economique, 38(3), 673-708.

[193] Sarno, L., Taylor, M. and Chowdhury, I. (2004). Nonlinear dynamics in deviations from the law of one price: a broad-based empirical study. Journal of International Money and Finance, 23(1), 1-25.

[194] Schotman, P. and van Dijk, H.K. (1991). A Bayesian analysis of the unit root in real exchange rates. Journal of Econometrics, 49(1-2), 195-238.

[195] Serletis, A. and Shahmoradi, A. (2007). Chaos, self-organized criticality, and SETAR nonlinearity: An analysis of purchasing power parity between Canada and the United States. Chaos, Solitons & Fractals, 33(5), 1437-44.

[196] Sims, C.A. and Uhlig, H. (1991). Understanding Unit Rooters: A Helicopter Tour. Econometrica, 59(6), 1591-9.

[197] Sollis, R., Leybourne, S.J. and Newbold, P. (2000). Stochastic Unit Roots Modelling of Stock Price Indices. Applied Financial Economics, 10(3), 311-25.

[198] Sowell, F. (1992/0). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics, 53(1-3), 165-88.

[199] Stock, J.H. and Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. Journal of Forecasting, 23(6), 405-30.

[200] Stock, J.H. and Watson, M. (1998). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. NBER Working Papers no. 6607.

[201] Svensson, L. (2002). The Inflation Forecast and the Loss Function. Discussion Papers in its Series CEPR Discussion Papers with Number 3365.

[202] Tanner, M.A. and Wong, W.H. (1987). The Calculation of Posterior Distributions by Data Augmentation: Rejoinder. Journal of the American Statistical Association, 82(398), 548-50.

[203] Tay, A.S. and Wallis, K.F. (2000). Density forecasting: A survey. Journal of Forecasting, 19, 235-54.

[204] Taylor, A.M.R. and van Dijk, D. (2002). Can Tests for Stochastic Unit Roots Provide Useful Portmanteau Tests for Persistence? Oxford Bulletin of Economics and Statistics, 64, 381-97.

[205] Taylor, M.P. (1995). The Economics of Exchange Rates. Journal of Economic Literature, 33(1), 13-47.

[206] Taylor, M.P., Peel, D.A. and Sarno, L. (2001). Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles. International Economic Review, 42(4), 1015-42.

[207] Taylor, M. and Sarno, L. (1998). The behavior of real exchange rates during the post-Bretton Woods period. Journal of International Economics, 46(2), 281-312.

[208] van Dijk, D., Teravirta, T. and Philip, H.F. (2002). Smooth Transition Autoregressive Models–A Survey of Recent Developments. Econometric Reviews, 21(1), 1-47.

[209] 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(430), 614-8.

[210] Wago, H. (2004). Bayesian estimation of smooth transition GARCH model using Gibbs sampling. Mathematics and Computers in Simulation, 64(1), 63-78.

[211] Wallis, K.F. (1995). Large-Scale Macroeconometric Modeling. In Pesaran, M.H. andWickens, M.R. (eds.), Handbook of Applied Econometrics. Oxford: Blackwell. 312-55.

[212] Wallis, K.F. (2003). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. International Journal of Forecasting, 19(2), 165-75.

[213] Watson, M. and Stock, J.H. (2007). Why has inflation become harder to forecast? Journal of. Money Credit and Banking, 39(1), 3-33.

[214] Weiss, A.A. (1986). ARCH and Bilinear Time Series Models: Comparison and Combination. Journal of Business & Economic Statistics, 4(1), 59-70.

[215] West, K.D. (1996). Asymptotic Inference about Predictive Ability. Econometrica, 64(5), 1067-84.

[216] West, M. and Harrison, J. (1997). Bayesian forecasting and dynamic models. 2nd ed. New York ; London: Springer.

[217] Yoon, G. (2006). A note on some properties of STUR processes. Oxford Bulletin of Economics & Statistics, 68(2), 253-60.

[218] Yoon, G. (2005). Stochastic Unit Roots in the Capital Asset Pricing Model? Bulletin of Economic Research, 57(4), 3307-78.

[219] Zellner, A. (1971). An introduction to Bayesian inference in econometrics. New York ; Chichester: John Wiley.