View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

University of Leicester

DEPARTMENT OF ECONOMICS

BAYESIAN ANALYSIS OF DETERMINISTIC TIME TREND AND CHANGES IN PERSISTENCE USING A GENERALISED STOCHASTIC UNIT ROOT MODEL

Fuyu Yang, University of Leicester, UK

Working Paper No. 07/11
October 2007

Bayesian Analysis of Determinisitic Time Trend and Changes in Persistence Using a Generalised Stochastic Unit Root Model [1]

Fuyu Yang*[2]
* *University of Leicester, Department of Economics*

Version: March, 2007

This paper makes use of the novel Generalized Stochastic Unit Root (GSTUR) model, Bayesian model estimation and model comparison techniques 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 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. Empirical analysis indicates that the GSTUR model could provide new insights on time series studies.

Key Words: Stochastic Unit Root, MCMC, Bayesian

1. INTRODUCTION

Application of econometric tests indicates that many macroeconomic time series contain unit roots and are therefore nonstationary I(1) processes. Some of these results are in contradiction with economic theories, such as the Purchasing Power Parity, which imply the stationarity of some series. Further development in some 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. This paper makes use of Bayesian techniques in testing for deterministic time trend and changes in persistence in time series with a parameter nonlinear model, Generalised Stochastic Unit Root (GSTUR) model

$$\nu_t = y_t - \delta t - \gamma \tag{1}$$

$$\nu_t = \exp(\alpha_t)\nu_{t-1} + \sum_{i=1}^{l} \lambda_i \triangle \nu_{t-i} + \varepsilon_t \tag{2}$$

$$\alpha_t = \phi_0 + \phi_1 \alpha_{t-1} + \cdots + \phi_p \alpha_{t-p} + \eta_t \tag{3}$$

where ε_t is $i.i.d.N(0, \sigma^2_\varepsilon)$ and η_t is $i.i.d.N\left(0, \sigma^2_\eta\right)$. While the original STUR model,

$$y_t = \exp(\alpha_t) y_{t-1} + \varepsilon_t \tag{4}$$

$$\alpha_t = \phi_0 + \phi_1 \alpha_{t-1} + \eta_t \tag{5}$$

[1] I would like to thank Roberto Gonzalez, Rodney Strachan, Stephen Hall and Wojciech Charemza for many useful suggestions.

[2] Corresponding author: Tel: 0044-116-252-5338. Email address: fay1@le.ac.uk

where ε_t is $i.i.d.N(0, \sigma_\varepsilon^2)$ and η_t is $i.i.d.N\left(0, \sigma_\eta^2\right)$, is proposed by Granger and Swanson (1997).

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 series that appear to be nonstationary after differencing or detrending. As evidence were found for changes in persistence with applications of U.S. macroeconomic data sets (Kim 2000, Kim 2002, Busetti and Taylor 2004, Harvey et al. 2006), it is sensible to argue that a deterministic time trend hypothesis is often rejected because of the changes in persistence. Therefore, there is always uncertainty as to whether a macroeconomic series is trend stationary (TS) or difference stationary (DS) or neither (see Newbold et al. 2001). The crucial questions are: 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. Standard algebra shows that the GSTUR model has the following desirable features: 1. the change in persistence could be different at any time point. 2. The parameter characterizing a new degree of persistence 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, estimation involved was problematic. One motivation of using Bayesian techniques is that estimations for this highly parameterized model can only be achieved by Markov Chain Monte Carlo (MCMC) techniques. Granger and Swanson (1997) used two methods to estimate the parameters in a STUR model (Equation 4,5), which produced 'wild estimates' that are 'fairly imprecise'. Simulations via MCMC techniques could not only shed light on the distribution properties for any feature of interest, but also be diagnosed (see Carlin and Chib 1995, Geweke 1989). Jones and Marriott (1999) provided a Bayesian method for parameter estimations for the STUR model. In this paper, Bayesian analyses of a GSTUR model are presented. From 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 is the proposition favoured according to the data information.

The remainder of the paper is organized as follows. Section 2 presents the GSTUR models, estimation and model selection in a Bayesian framework together with experiments using simulated data series. Section 3 presents the empirical results with applications of Nelson and Plosser's S&P 500 series and the U.K. /U.S. long run exchange rates using a GSTUR model. Section 4 contains brief concluding remarks.

2. BAYESIAN INFERENCES

According to the GSTUR model (Equation 1, 2 and 3), α_t is literally a $AR(p)$ process. The roots of the polynomials α_t are restricted within the unit circle. Then an unconditional mean μ_α of the stationary process α_t has the following expressions:

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

Also,

$$\rho_t = \exp(\alpha_t) \tag{6}$$

We begin by introducing some notations: F_t denotes the history of observations sequence up to time t, $F_t = (y_1, \cdots, y_t)'$, y denotes the whole sample of observations with a sample size of N, $y = (y_1, y_2, \cdots, y_n)'$. Latent variable α, where $\alpha = (\alpha_{1-p}, \cdots, \alpha_0, \alpha_1, \cdots, \alpha_{T-1}, \alpha_T)'$ [3], denotes a series of stochastic roots over the time T period, which are structurally unobservable and indicate changes in the persistence of y. Initial values is $\alpha_{initial} = (\alpha_{1-p}, \cdots, \alpha_0)'$.Vector ϕ and λ are defined as $\phi = (\phi_1, \cdots, \phi_p)'$, $\lambda = (\lambda_1, \cdots, \lambda_l)'$.

According to equation (6), ρ_t then is a vector $\left(\rho_{1-p}, \cdots, \rho_0, \rho_1, \cdots, \rho_{T-1}, \rho_T\right)'$ associate with α_t, $t = (1-p, \cdots, T)$. The error precisions are denoted as $h_\varepsilon = \sigma_\varepsilon^{-2}$ and $h_\eta = \sigma_\eta^{-2}$. $\theta = \left(\gamma, \delta, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2\right)$ stands for a vector of all the parameters of interest. ρ_t varies stochastically in the GSTUR process. To investigate if a process undergoes shifts in persistence, or being parameter nonlinear, we can make inferences from the estimates of ρ_t. Process may maintains stationarity (I(0)) if $\rho_t < 1$, but becomes a process of higher persistence (I(1)) if $\rho_t > 1$. This idiosyncratic property of a STUR process makes it very difficult to distinguish stochastic process from a Random Walk (RW)[4] process. The difference stationary RW model is nested within the GSTUR model at the point where $\mu_\alpha = 0$ and $\sigma_\eta^2 = 0$ such that ρ_t will be a constant and equals to 1. Therefore, the behaviours of a process could be learnt about by observing the distribution of μ_α and the evolution of ρ_t over time. Hence, our focus of estimations is on μ_α, ρ_t and δ.

2.1. Bayesian Model Estimation

The nature of Bayesian inference lies in the prior beliefs of various possible hypotheses and the data information. Base on the Bayes' theorem,

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

the original belief, the 'prior' $p\left(M \mid I_0\right)$, could be adapted to the 'posterior' $p\left(M \mid y, I_0\right)$ taking the data information $p\left(y \mid M\right)$ into account. In the context of Bayesian framework, estimations via simulation is virtually revolutionized by the Markov Chain Monte Carlo (MCMC) method (Gelfand and Smith 1990, Tanner and Wong 1987). With MCMC approach to estimate, we are able to plot the whole density distribution for the parameter of interest. The latent variable α is treated as 'pseudo' parameter whose behaviour is governed by values of the elements in θ. As α is stochastic and unobserved that it must be estimated as must the parameters in θ. Thus, Bayesian analysis uses the joint density of and conditional on y, will be

$$p\left(\theta, \alpha | y\right) \propto p\left(y | \alpha, \theta\right) p\left(\alpha, \theta\right)$$

The form of priors and generally follow the general recommendations of Jones and Marriott (1999) with some important exceptions. We assume that each parameter is a priori independent and formulate prior distributions on the parameters of interest. The notations of relevant distributions are given as follows: $f_N\left(\mu, \sigma^2\right)$denotes a normal distribution with a mean of μ and a variance of σ^2, $f_{MN}\left(M, V\right)$denotes a multivariate normal distribution with a mean of M and a variance covariance matrix of V, f_Γ denotes a Gamma distribution, and f_Γ^{-1}denotes an Inversed Gamma distribution.

The prior distributions are: $\mu_\alpha \sim i.i.df_N\left(\ln 0.9, 0.1^2\right)$, which pushes μ_α moving towards with a small variance; multivariate normal for $(\gamma, \delta)$and λ; $\phi_i \sim i.i.df_N\left(\mu_{\phi_i}, \sigma_{\phi_i}^2\right)$ $i = 1, \cdots, p$. As μ_α is the mean of stationary $AR(p)$ process, if $\phi\left(L\right) = 1 - \phi_1 L - \cdots -$

[3] We also defined α_{-t} as $\alpha_{-t} = (\alpha_{1-p}, ..., \alpha_{t-1}, \alpha_{t+1}, ..., \alpha_{T-1}, \alpha_T)'$

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

$\phi_p L^p$ is a polynomial of order p in the lag operator, the roots of equation $\phi(z) = 0$ should be all greater than one in absolute values. Thus, the prior density function of ϕ_i is chosen as $\phi_i \sim f_N(0,1)\,1(\|z_j\| > 1)$, where $1(A)$ is the indicator function for the event A that $\phi_i s$ are jointly truncated within the stationary region. The error precisions' prior are $h_\varepsilon \sim f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ and $h_\eta \sim f_{\Gamma\eta}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$, where $\underline{\alpha}_\varepsilon = \frac{1}{256}, \underline{\beta}_\varepsilon = 256$, $\underline{\alpha}_\eta = 1.5$ and $\underline{\beta}_\eta = 0.03$. Please refer to Appendix (A) for prior elicitations. Given the above prior information, the goal of the Bayesian analysis is to learn about the parameter from the augmented posterior distribution.

The joint posterior density for is then

$$
\begin{aligned}
p(\theta, \alpha \mid y) &\propto \prod_{t=2}^{N} p(y_t \mid \alpha, \theta, F_{t-1})\, p(\alpha, \theta) \\
&\propto \prod_{t=2}^{N} p(y_t \mid \alpha, \theta, F_{t-1})\, p(\alpha_t \mid \alpha_{-t}, \theta)\, p(\theta)
\end{aligned}
\tag{7}
$$

where F_{t-1}denotes the history of observations sequence up to time $t-1$, with $F_{t-1} = (y_1, \cdots, y_{t-1})'$.

Inference on the parameters can be conducted by producing sequential draws from this density by a MCMC procedure. Focusing on the posterior properties, Gibbs sampler is designed. A full analysis of posterior conditionals and the Codes (in Matlab) are available on reader's request. Jones and Marriott (1999) used the ratio of uniforms methods (see Devroye, 1986) to sample α_t. In this paper, the Metropolis-Hastings (M-H) algorithm is implemented to simulate α_t.

2.2. Bayesian Model Comparison

In this paper, the main concerns are the existence of a deterministic time trend (whether $\delta = 0$), the parameter nonlinearity in the process (whether ρ is time-invariant) and the value of l for the best model fit. A group of models holding different hypothesis 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 via Bayes Factors. The interested reader is directed to Kass and Raffrey (1995) who provide detailed discussion on issues concerning on Bayes factors. Bayes Factors for model comparison has advantage over-parameterizations (see Koop and Potter, 1999). According to Kass and Raftery (1995), the Bayes Factor B_{ij} (choice between model M_i and model M_j) is expressed as

$$
B_{ij} = \frac{p(y|M_i)}{p(y|M_j)} \tag{8}
$$

where $\frac{p(y|M_i)}{p(y|M_j)}$ is the marginal likelihood ratio between model M_i and M_j. The strength of evidence in favours of model M_i versus M_j is evaluated according to the log Bayes factor scale in Table (1) under Kass and Raffrey (1995)'s classifications:

The marginal likelihood of the GSTUR models can be approximated from the posterior samples using the approach introduced by Chib (1995), which is also reviewed by Han and Carlin (2001). According to Chib (1995), the marginal density of $y = (y_2, \cdots, y_n)'$can be written as

$$
p(y) = \frac{\prod_{t=2}^{N} p(y_t \mid \theta, F_{t-1}) p(\theta)}{p(\theta \mid y)}
$$

TABLE 1
Bayes factor scale comparing model i with model j

$log_{10}B_{ij}$	B_{ij}	Evidence against model j
$0-1$	$1-3$	Not worth more than a bare mention
$1-3$	$3-20$	Positive
$3-5$	$20-150$	Strong
>5	>150	Very Strong

with all integrating constants included[5]. As the above identity holds for any θ, say θ^*, the value of posterior density $p(\theta^* \mid y)$ can be estimated as $\widehat{p}(\theta^* \mid y)$ by using the Monte Carlo samples. Then, the log marginal likelihood can be approximated as:

$$\ln \widehat{p}(y) = \sum_{t=2}^{N} \ln p(y_t \mid \theta^*, F_{t-1}) + \ln p(\theta^*) - \ln \widehat{p}(\theta^* \mid y) \tag{9}$$

For a more accurately approximation, θ^*should be evaluated at a high density point, which are chosen as posterior means in this paper. To achieve $\ln p(\theta^*)$ and $\ln \widehat{p}(\theta^* \mid y)$ are straight forward. However, according to Equation (4), y_t is a function of latent variable α_t, thus $\ln p(y_t \mid \theta^*, F_{t-1})$ evolves calculation of

$$p\left(y_t \mid \theta^*, F_{t-1}\right) = \int p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right) p\left(\alpha_t \mid \theta^*, F_{t-1}\right) d\alpha_t \tag{10}$$

As α_t is non-observable, the exact integrals are hard to obtain. However, with the help of Monte Carlo averaging $p\left(y_t \mid \alpha_t, \theta^*, F_{t-1}\right)$ over the 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 $p\left(y_t \mid \theta^*, F_{t-1}\right)$ from the following:

$$p\left(y_t \mid \theta^*, F_{t-1}\right) \simeq \frac{1}{M} \sum_{g=1}^{M} p\left(y_t \mid \alpha_t^{(g)}, \theta^*, F_{t-1}\right) \tag{11}$$

However, a sample of $\alpha_t^1, \cdots, \alpha_t^M$ from $p\left(\alpha_t \mid \theta^*, F_{t-1}\right)$ would need a sample of $\alpha_{t-1}^1, \cdots, \alpha_{t-1}^M$ from the $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$ as

$$p\left(\alpha_t \mid \theta^*, F_{t-1}\right) = \int p\left(\alpha_t \mid \alpha_{t-1}, \theta^*, F_{t-1}\right) p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right) d\alpha_{t-1} \tag{12}$$

An Auxiliary Particle Filter (APF) method introduced in Pitt and Shephard (1999a) is applied here to get samples from $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$. The idea of APF is that if we can get a sample of $\alpha_{t-1}^1, \cdots, \alpha_{t-1}^G : M \le G$ from $p\left(\alpha_{t-1} \mid \theta^*, F_{t-2}\right)$, when the new data y_{t-1} arrives and data information is updated from F_{t-2} to F_{t-1}, we can get a sample of $\alpha_{t-1}^1, \cdots, \alpha_{t-1}^M$ according to the weights of likelihood $p\left(y_{t-1} \mid \alpha_{t-1}, \theta^*, F_{t-2}\right)$. Then, the resampled $\alpha_{t-1}^1, \cdots, \alpha_{t-1}^M$ according to the likelihood weights $p\left(y_{t-1} \mid \alpha_{t-1}, \theta^*, F_{t-2}\right)$ will approximate to a sample from $p\left(\alpha_{t-1} \mid \theta^*, F_{t-1}\right)$. The algorithm is explained as follows:

Algorithm: Estimate the log Likelihood for the marginal likelihood using 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 initial M values

[5] ϕ is truncated to satisfy the stationary restriction in the transition equation. The integrating constant for prior ϕ can be evaluated in a simulation manner. For details, please refer to Judge et.al (1985, p128).

of $\underline{\alpha}_2^{(g)}: g = 1, \cdots, M$ can be set as a $M \times p$ zero matrix or a sample of M draws $(\underline{\alpha}_2^{(1)}, \cdots, \underline{\alpha}_2^{(M)})'$ of $\underline{\alpha}_1$ from the conditional prior $p(\underline{\alpha}_2|\theta^*)$.

1. Fix $t = 2$.
   (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 $p(y_t|\theta^*)$ is given by:

   $$\hat{p}(y_t|\theta^*) = \frac{1}{M}\sum_{g=1}^{M} p(y_t|\theta^*, \alpha_t^{\triangle(g)}, F_{t-1}) \tag{13}$$

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:

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

   Get R draws $(k_1, \cdots, k_R)$ from the discrete distribution defined on the integers $(1, 2, \cdots, M)$ with probabilities $\overline{w}_1, \cdots, \overline{w}_M$. Note that each value of k_r is used to indicate a value of $\underline{\alpha}_t^{(k_r)}$ (and of $\hat{\alpha}_t^{k_r}$)
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

   $$\alpha_t^{*(r)} \sim f_N(\underline{\alpha}_t^{k_r}\phi, \sigma_\eta^{2*}) \tag{14}$$

   Note that $\left(\alpha_t^{*(r)} : r = 1, \cdots, R\right)$ is a sample from $p(\alpha_t|\theta^*, F_{t-1}, \underline{\alpha}_t)$
4. Resample the $R \times 1$ vector $\left(\alpha_t^{*(r)} : r = 1, \cdots, R\right)'$ M times with probabilities $\overline{w}_r^*$ defined as:

   $$\begin{aligned} w_r^* &= \frac{p(y_t|\theta^*, F_{t-1}, \alpha_t^{*r})}{p(y_t|\theta^*, F_{t-1}, \hat{\alpha}_t^{k_r})} \qquad (15) \\ \overline{w}_r^* &= \frac{w_r^*}{\sum_{r=1}^{M} w_r^*} \end{aligned}$$

   Then, the resampled $M \times 1$ vector, which contains values $(\alpha_t^{(1)}, \cdots, \alpha_t^{(M)})'$ is (approximately) distributed as $p(\alpha_t|\theta^*, F_t, \underline{\alpha}_t)$. This $(\alpha_t^{(1)}, \cdots, \alpha_t^{(M)})'$ stack 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. For each $\underline{\alpha}_{t+1}^g : g = 1, \cdots, M$, sample a value $\alpha_{t+1}^{\triangle(g)}$ using the transition density:

   $$\alpha_{t+1}^{\triangle(g)} \sim f_N(\underline{\alpha}_{t+1}^g\phi, \sigma_\eta^{2*})$$

   Note that $\alpha_{t+1}^{\triangle(g)} : g = 1, \cdots, M$ is (approximately) a sample from $p(\alpha_{t+1}|\theta^*, F_t, \underline{\alpha}_{t+1})$.

6. Fix $t = t + 1$ until $t = N$. An estimate of $p(y_{t+1}|\theta^*, F_t)$ is given by[6]:

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

7. Go to step 2. Finally, the estimate of the log likelihood is

$$\log \widehat{p}\,(y \mid \theta^*) = \sum_{t=2}^{N} \log \widehat{p}\,(y_t \mid \theta^*, F_{t-1}) \tag{16}$$

A particle filtering method could recursively deliver sequence of draws of $p\left(\alpha_2^{(g)} \mid \theta^*, F_2\right)$, ..., $p\left(\alpha_t^{(g)} \mid \theta^*, F_t\right)$,..., $p\left(\alpha_n^{(g)} \mid \theta^*, F_n\right)$. Note that there other less efficient methods to calculate the likelihood ordinate $p(y|\theta^*)$. For example, one could get random draws of $(\alpha_1, \alpha_2, ..., \alpha_n|\theta^*)$ from the conditional prior $p(\alpha_1, \alpha_2, ..., \alpha_n|\theta^*)$ and average the conditional likelihood $p(y|\theta, \alpha_1, \alpha_2, ..., \alpha_n)$ over these values. Even though this would deliver also an estimate of the likelihood $p(y|\theta^*)$, it is less efficient than using expression (15). This is because many of the draws from the prior will be in the region where the conditional likelihood $(p(y|\theta, \alpha_1, \alpha_2, ..., \alpha_n))$ is very small, and would therefore contribute little to the accuracy of the estimate. In contrast, the method we are using draws the α values using the data information, over a region where the conditional likelihood has greater weight. Therefore, we could say that the other method is 'blind', because it ignores the data information.

With the marginal likelihood identity from equation (9), estimated posterior ordinates using a reduce Gibbs runs and the estimated likelihood from an Auxiliary Particle Filter from equation (16), we are able to produce marginal likelihood of any model of interest for model comparison purpose.

As the models of interest are all nested within the GSTUR model, we can employ techniques that take advantage of this feature to estimate the marginal likelihood. Although the Random Walk model also nests within the GSTUR model, it is computational simpler to compute the marginal likelihood directly. As in this case, the marginal likelihood can be obtained analytically as follows:

$$p_{RW}(y) = \int p(y|\sigma_\varepsilon^2)p(\sigma_\varepsilon^2)d\sigma_\varepsilon^2 \tag{17}$$

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

2.3. Evaluation of the MCMC Using Artificial Data

To evaluate the estimation efficiency, a series of artificial data with a sample size of 118 is simulated from the GSTUR Process with parameters specified as the following:

$$\nu_t = y_t - 1 - 0.05t \tag{18}$$

$$\nu_t = \exp(\alpha_t)\nu_{t-1} + 0.8 \triangle \nu_{t-1} + \varepsilon_t$$

[6] Here I use this expression that I am more familiar with. In your chapter you give an expression based on weights, which may also be correct. If you use the expression with weights, I recommend that you put a reference.

$$\alpha_t = -0.03125 + 0.8\alpha_{t-1} + \eta_t$$

where $\varepsilon_t \sim i.i.dN(0,1)$ and $\eta_t \sim i.i.dN(0, 0.01^2)$. $\mu_\alpha = -0.15625$

The graph of the simulated data is plotted in Figure (1):

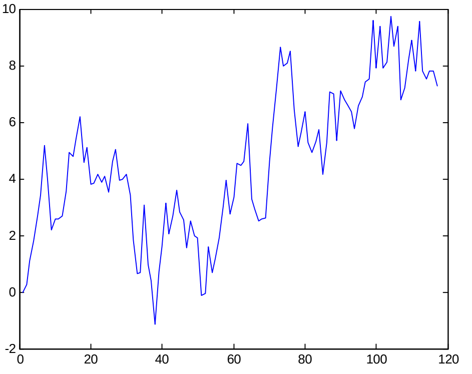


Fig 1. Plot of Simulated Data

2.3.1. MCMC Estimation Efficiency

A GSTUR model with a constant, a trend, $l = 1$ and $p = 1$ is estimated by running the Gibbs sampler for $8,000$ replications with the initial $2,000$ discarded. Results together with the 'true' values are provided in Table (2) for the purpose of evaluating the efficiency of the estimates. If the absolute value of CD is smaller than 1.96, convergence of the MCMC is achieved. From the Table (2), the Gibbs Sampler converged. To illustrate the MCMC convergence speed, the estimates of are taken for examples. Figure 2 (a-b), 3 (a-b), and 4(a-b) plot the actual value and the histogram for the first 50, 200, and $6,000$ (after burn-in) draws of from the Gibbs Sampler respectively. The draws of quickly converge to the 'true' value in all the experiments. With more replication time, the histogram exhibits better normality. Figure (5-8) plots the histogram of the retained draws $6,000$ draws of γ, δ, $\sigma_\varepsilon^2$and σ_η^2. Under the controlled setting, the Gibbs sampler is able to provide efficient and accurate estimates.

TABLE 2
Estimated Results with Artificial Dataset

parameters	*trueV*	*Est.mean*	*s.t.d*	*CD*
γ	1	1.6960	0.5425	-0.2194
δ	0.05	0.0531	0.0059	0.1783
λ_1	0.8	0.9310	0.0477	0.4036
μ_α	-0.1563	-0.1356	0.0815	0.4263
ϕ_1	0.8	0.0507	0.2710	0.2112
σ_ε^2	1	0.7392	0.1234	-0.2337
σ_η^2	0.01^2	0.0935	0.0653	-0.0831

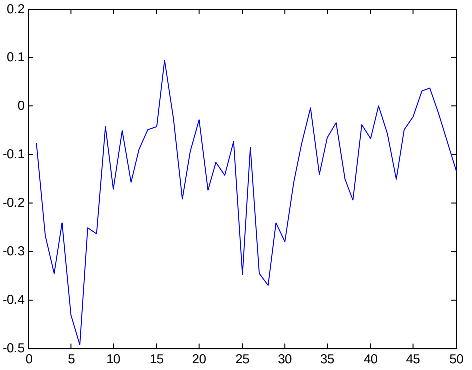


Fig 2 (a) Plot of first 50 μ_α draws

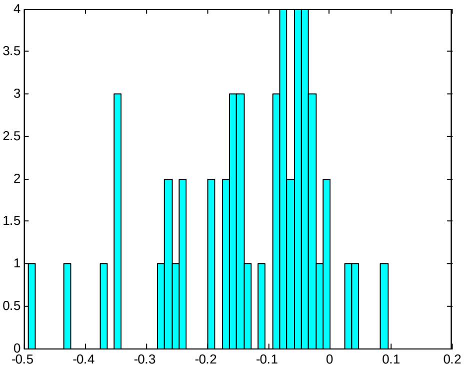


Fig 2 (b) Histogram of first 50 μ_α draws

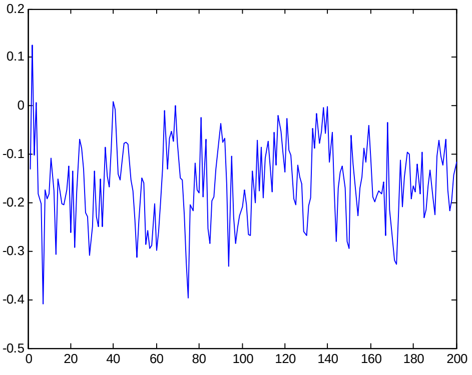


Fig 3 (a) Plot of first 200 μ_α draws

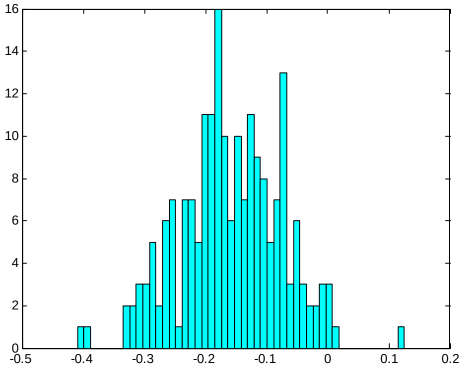


Fig 3 (b) Histogram of first 200 μ_α draws

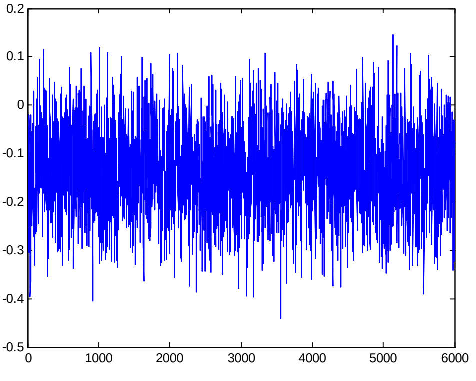


Fig 4 (a) Plot of 6,000 Actual Draws for μ_α

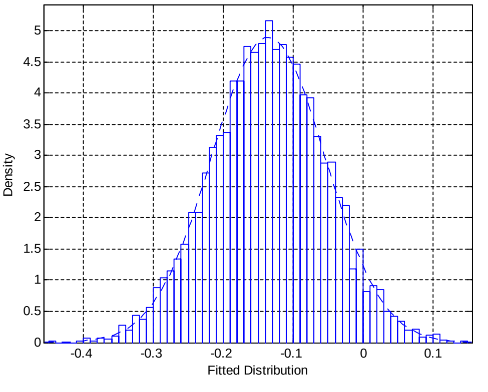


Fig 4 (b) Histogram of 6,000 μ_α draws after burn-in

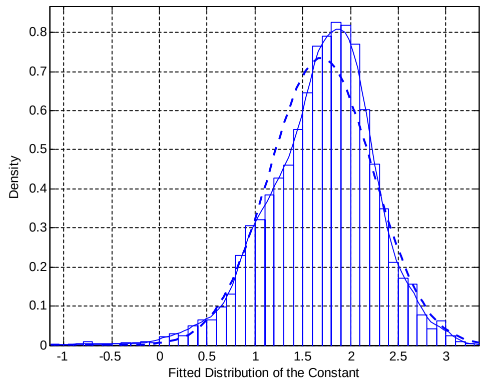


Fig 5 Histogram of 6,000 γ draws after burn-in

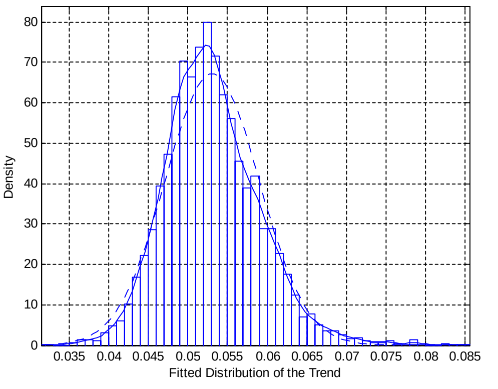


Fig 6 Histrogram of 6,000 Draws for δ

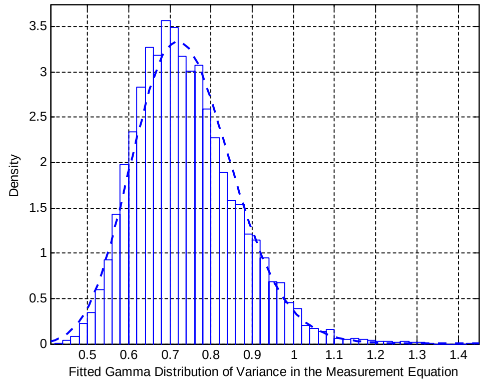


Fig 7 Histrogram of the 6,000 Draws for σ_{ε}^2

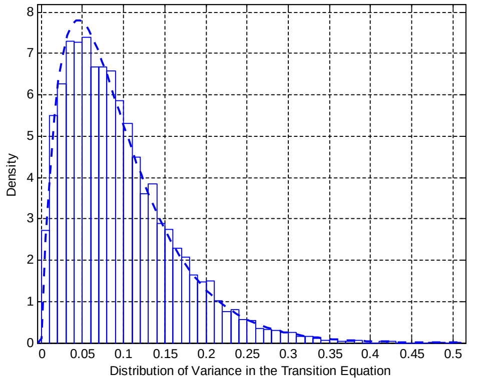


Fig 8 Histrogram of the 6,000 Draws for σ_{η}^2

2.3.2. Model Uncertainties

Under the controlled settings, we are able to reflect on the model uncertainties from the log marginal likelihood $\log [y \mid M]$. In this experiment, we would expect a significant support for a model with a constant, a trend, and nonlinearity in Stochastic Root specifications, since we know what the true DGP is. The marginal likelihoods of 15 models, which are specified as different combinations of $\left[\sum_{i=1}^{l} \lambda_i, \gamma, \delta\right]$, with $l = 0, 1, 2, 3, 4$ are evaluated.

Table (2) presents the marginal likelihood of the 15 models. The model with $\gamma \neq 0$, $\delta \neq 0$, and $l = 1$ is highly favoured, while the 'true' data is actually generated with these features. An Inversed Gamma $f_{\Gamma}^{-1}\left(\underline{\alpha}_{\varepsilon}, \underline{\beta}_{\varepsilon}\right)$ is selected as the prior of $p\left(\sigma_{\varepsilon}^2\right)$ for the RW model. Figure (9) plots the of the RW model with different values of $\left[\underline{\alpha}_{\varepsilon}, \underline{\beta}_{\varepsilon}\right]$, where $\underline{\alpha}_{\varepsilon}$ and $\underline{\beta}_{\varepsilon}$ vary from 0.01 to 5. We may see that the marginal likelihood of RW $\log [y \mid M_{RW}]$ maximized at -176. If we use the same prior parameters as in the GSTUR model, the RW model's marginal likelihood is only -271. Thus, under the same priors of $p\left(\sigma_{\varepsilon}^2\right)$, the nonlinear GSTUR model with $\gamma \neq 0$, $\delta \neq 0$ is more favoured than a RW model. In this case, the Bayes Factor is able to detect the nonlinearity in the data generating process.

Table (3) presents the marginal likelihood of the 15 models.

TABLE 3
Model Comparison Results under Controlled Settings

$\log[y \mid M]$			
lag length	$\gamma \neq 0, \delta \neq 0$	$\gamma \neq 0, \delta = 0$	$\gamma = 0, \delta = 0$
4	-211.6691	-240.3003	-339.3462
3	-204.2667	-234.2754	-331.8269
2	-195.6707	-224.7850	-323.9709
1	-190.2175^{*}	-218.6692	-314.6453
0	-245.7978	-248.8422	-324.9754

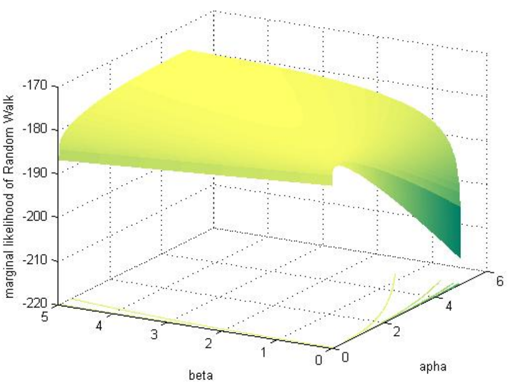


Fig 9 Surface plot of the Marginal Likelihood
with Random Walk Model

Koop (1999) state that when comparing nonlinear models with linear models, one obvious advantage of using Bayes Factor is that according to 'Occam's Razor', nonlinear models will be preferred only when expected nonlinearity does exist. The experimental results reflect on the model uncertainties and possible existence of the nonlinearities in the underlying process of time series.

3. EMPIRICAL ILLUSTRATIONS WITH A GENERALISED STUR MODEL

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&P500) indicates that a deterministic time trend might exists in the S&P500 series. A further application to U.K. /U.S. long-run exchange rate indicates the changing persistence consist with monetary events.

3.1. Empirical Results with Stock Price

S&P 500 yearly data set is chosen from the extended Nelson and Plosser's data set (from 1877 to 1988) for empirical application. This data set has been previously tested for an exact Unit Root, deterministic time trend and changing persistence (see Nelson and Plosser (1982), Kwiatkowski and Phillips et.al (1991), Gil-Alana and Robinson (1997)).

This data set has also been applied by Jones and Marriott (1999) with the original stochastic unit root model (Equation 4 and 5). In this paper, not only the S&P 500 data is applied with a GSTUR model for estimations, but also the model probabilities are evaluated to shed light on the model uncertainties.

3.1.1. Estimation and Efficiency

To ensure that the effect of the starting values in the MCMC algorithms is insignificant, we take 25,000 draws with the first 5,000 discarded. The correlogram plots serial correlations of the draws from the MCMC algorithm. Figures 10 (a-b) indicate that, for all the parameters 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. According to the CD from Table (4), the Gibbs sampler converges for all 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 Roots.

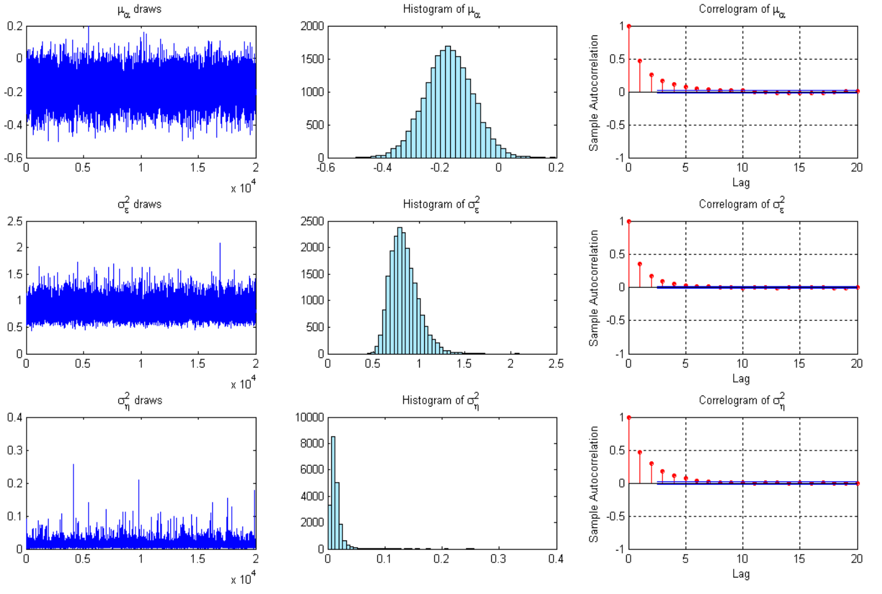


Fig 10(a) Gibbs Sampler for the SP500 Series for μ_α, σ_ε^2 and σ_η^2

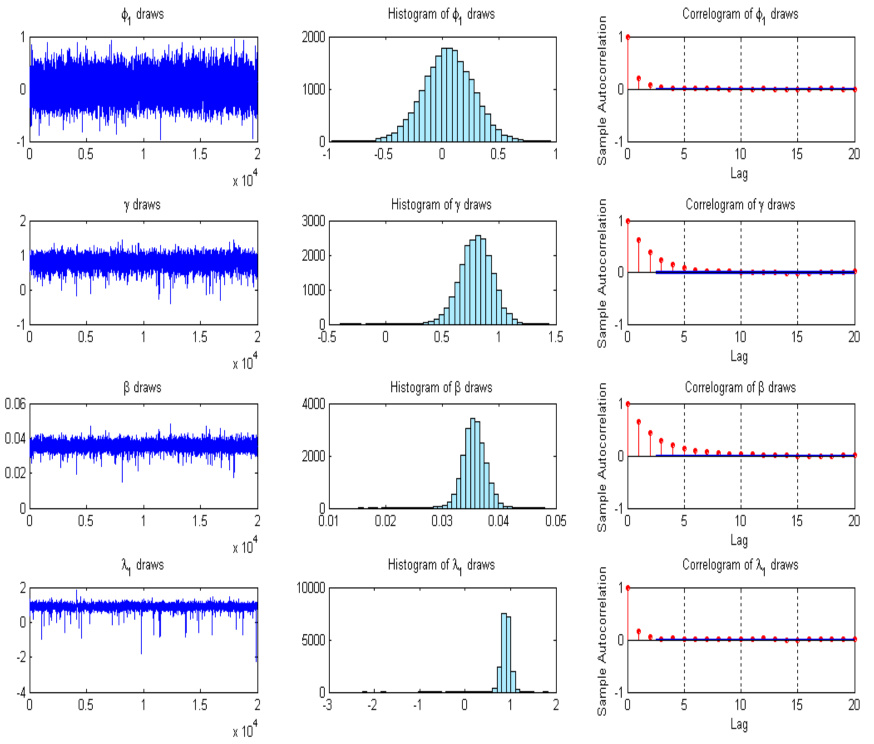


Fig 10(b) Gibbs Sampler for the SP500 Series ϕ_1,γ, β, and λ_1

TABLE 4
Estimations and Sampler Efficiency: The Generalized STUR with An Application of SP500 Series

	Prior		*Posterior*					
	Mean	*St.Dev*	*Mean*	*St.Dev*	*CD*	*Median*	95% *Posterior Band*	
μ_α	$\ln 0.9$	0.1	-0.1769	0.0842	-0.4232	-0.1770	-0.3151	-0.0385
σ_η^2	-	-	0.0136	0.0095	0.5695	0.0117	0.0043	0.0282
σ_ε^2	-	-	0.8368	0.1512	0.1166	0.8188	0.6256	1.1113
ϕ_1	†	†	0.0525	0.2208	-0.6449	0.0527	-0.3138	0.4130
γ	0	10^6	0.7938	0.1493	-0.0945	0.7995	0.5506	1.0221
β	0	10^6	0.0356	0.0022	0.0725	0.0356	0.0324	0.0389
λ_1	0	10^6	0.8968	0.1149	0.7400	0.8994	0.7394	1.0555

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

$-$: see Appendix A for description

3.1.2. Illustrations on Stochastic Unit Roots

To illustrate possible changes of the persistence in the underlying process[7], the estimated roots α_t (from 1877-1988) from the Gibbs Sampler are plotted. With different model specifications, the estimates of the roots are quite different. From Figure 11 (a-e), with different model specifications, the roots of the Stock prices series vary around $E\left[\exp\left(\mu_\alpha\right)\right]$ in the stationary region for most of the time, but go beyond 1 at certain time points. At these time points, there might be changes in the persistence of the underlying process. These indicate that the STUR process may not be easily distinguished from a linear unit root process, which may explains the evidence of a Unit Root in the Stock price series (Nelson and Plosser 1982, Kwiatkowski and Phillips 1991).

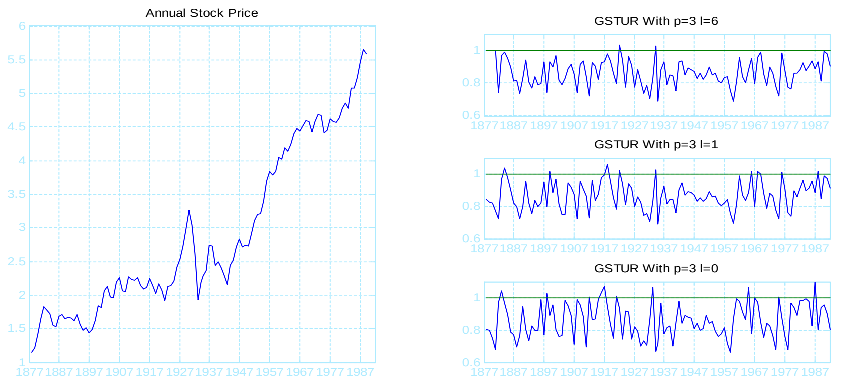


Fig 11(a). Unrestricted GSTUR model with $p = 3$

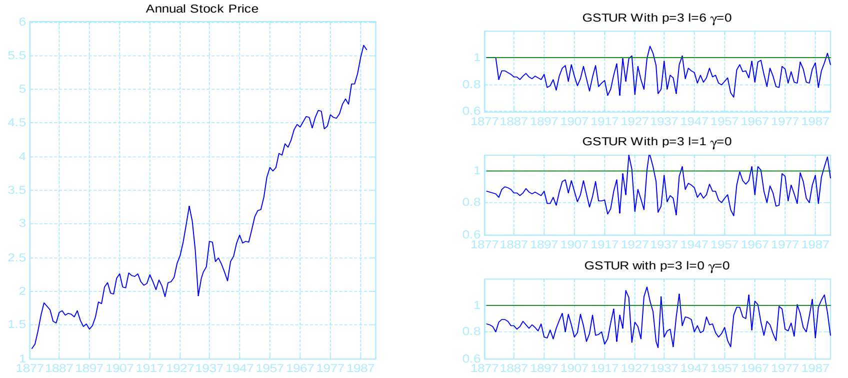


Fig 11(b). GSTUR Model with a Deterministic Time Trend $p = 3$

[7] In other words, the integration degrees of the series shift from I(0) to I(1), or vice versa.

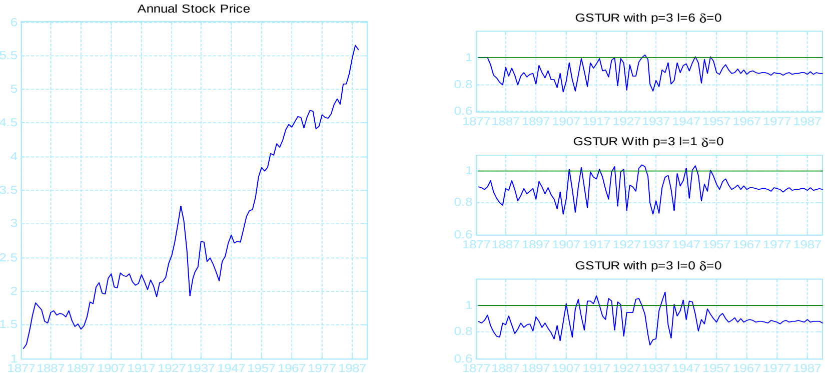


Fig 11(c). GSTUR Model with a Drift $p = 3$

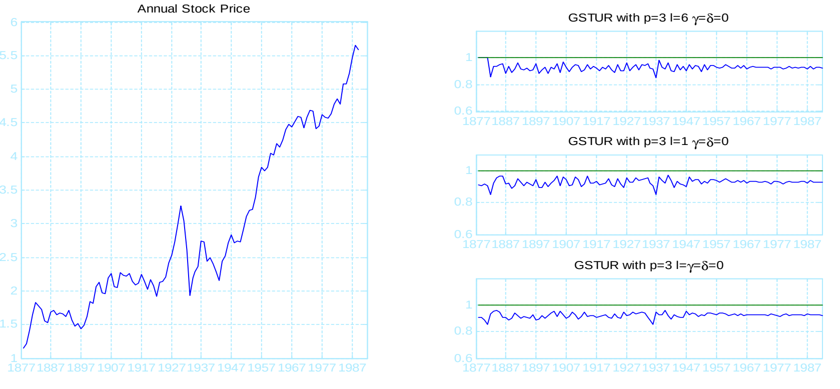


Fig 11(d). GSTUR with No Drift or Deterministic Time Trend $p = 3$

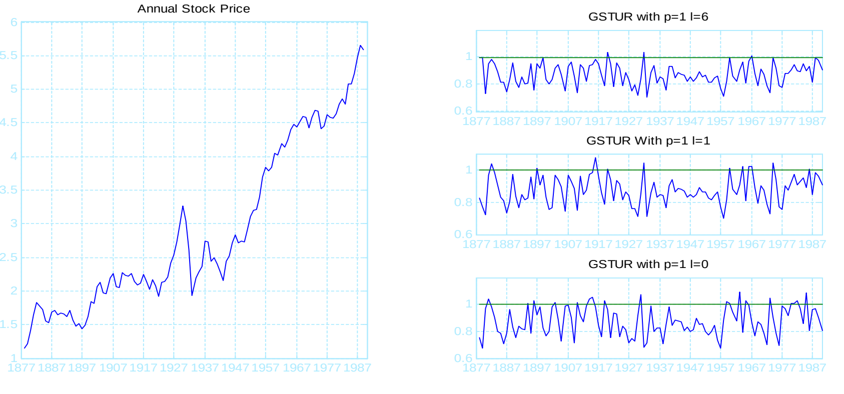


Fig 11(e). Unrestricted GSTUR with $p = 1$

3.1.3. Model Selection Results

From Figure 11(a-e), imposing $\delta = 0$ or $\gamma = \delta = 0$ change the results significantly. Koop (1994) points that imposing restrictions on the deterministic time trend is ruling out the possibility of a deterministic trend so that any trend behaviour 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 best fitting model. From Table (5), with the same lengths of l and p, the unrestricted GSTUR models are the most favoured models according to the highest marginal likelihoods. Different choices of p do not affect the log marginal likelihood $\log[y \mid M]$ much. However, the length l determines the fits of the model. From Figure (3.1.3), with the same prior as in the GSTUR model, the log marginal likelihood of a RW model is only -261.2901. Thus, an unrestricted GSTUR model with $l = 1$, $p = 1$ is strongly preferred over a RW model under the same choice of prior $p\left(\sigma_\varepsilon^2\right)$. Taking the model uncertainties into account by Bayesian Model Averaging, if we just consider the RW model and the unrestricted GSTUR models (with $l = 1$), the sample series has a 99% probability of being a stochastic unit root process[8].

TABLE 5

Marginal Likelihoods for Different Model Specifications

	unrestricted	$\gamma = 0$	$\beta = 0$	$\gamma = \beta = 0$
$(a)\, p = 3$				
$l = 0$	-139.9040	-87.2875	-113.3235	-256.0482
$l = 1$	17.3340	-18.2090	-83.4698	-258.2952
$l = 2$	11.6699	-39.3168	-87.2765	-265.0540
$l = 3$	-20.3548	-52.3195	-104.0309	-262.5562
$l = 4$	-41.175	-81.3793	-111.0549	-294.1133
$(b)\, p = 2$				
$l = 0$	-137.0078	-87.9067	-103.8562	-252.0244
$l = 1$	18.9501	-19.3308	-82.8389	-260.0658
$l = 2$	8.5642	-46.2332	-87.8234	-261.7946
$l = 3$	-12.1238	-48.6490	-97.5760	-267.1735
$l = 4$	-25.2809	-55.0828	-107.6392	-273.3519
$(c)\, p = 1$				
$l = 0$	-138.1189	-88.3923	-110.6433	-241.3139
$l = 1$	20.3223	-8.4980	-83.2075	-250.8019
$l = 2$	11.8530	-48.3653	-86.4118	-260.8656
$l = 3$	-9.7283	-47.3286	-101.5861	-271.6706
$l = 4$	-51.2533	-53.2768	-106.1881	-273.1563

[8] The weight averaged model $\overline{M}$ can be expressed as $\overline{M} = M_i p(M_i) + M_j p(M_j)$, where $p(M)$ indicates the model probabilties. If we just consider two models M_i and M_j, $p(M_j) = \frac{p(y|M_j)}{p(y|M_j)+p(y|M_i)} = \frac{1}{1+Bij}$

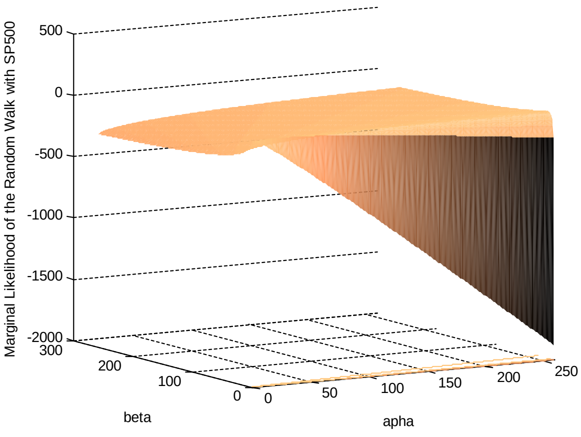


Fig 12 Surface plot of the log Marginal Likelihood for a Random Walk model with SP500

3.2. Empirical Results with Long-run Real Exchange Rate

Another empirical application for the GSTUR model is to analyze the monthly U.K. /U.S. real exchange rates from Jan 1885 to Feb 1995. This data was tested for a unit root with the Augmented Dickey-Fuller test by Engel and Kim (1999). The unit root was rejected at the 5% level. In this section, a restricted GSTUR model (setting $\gamma = \delta = 0$) is applied for the analysis. 25,000 iterations were taken with the first 5,000 discarded. Table (6) reports the estimation results.

Figure (3.2) plots the U.K. /U.S. real exchange rates, nominal exchange rates and estimated roots. For a review of the historical monetary events within the 111-year span, please refer to the summaries in Engel and Kim (1999). From Figure (3.2), the U.K. /U.S. long run real exchange rate is highly persistent according to the estimated Stochastic Roots. The range of the roots is narrow (from 0.98-1.015) and the roots are below 1 for most of the time. At certain time points, the roots jump to or above 1, which are marked with †. At these time points, the series can be recognized as nonstationary and/or explosive. We find that a change of the persistence in the series normally goes with an important monetary event.

TABLE 6

Estimations and Sampler Efficiency: The Generalized STUR with An Application of 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	−0.0211	0.0510	0.0204	−0.0090	−0.1330	0.0276
σ^2_η	−	−	0.0002	0.0002	−0.1093	0.0001	0.0000	0.0006
σ^2_ε	−	−	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

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

− : see Appendix A for description

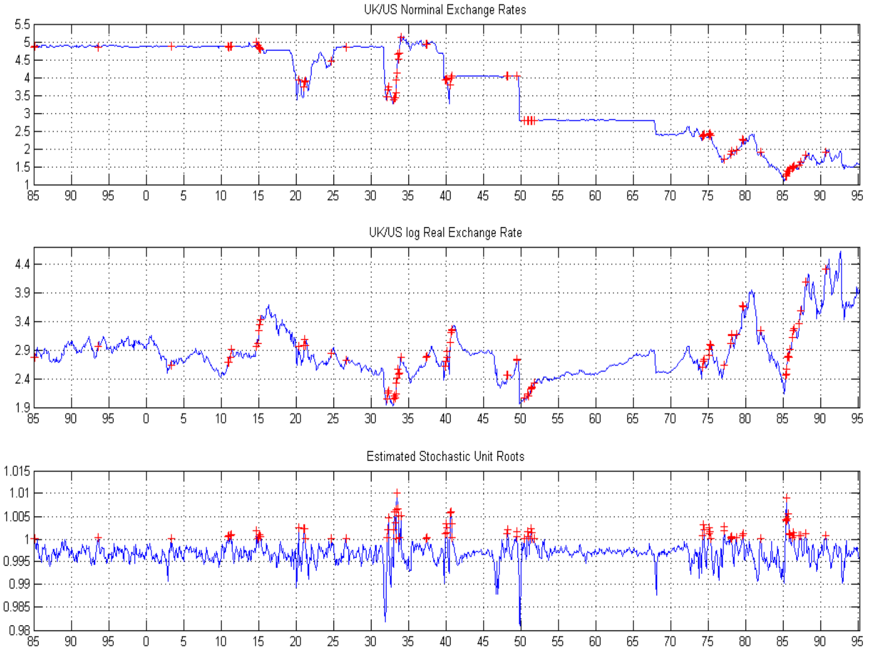


Fig 13. U.K./U.S. Long Run Exchange Rates and Estimated Stochastic Unit Roots

4. CONCLUSIONS

A coefficient nonlinear model, the GSTUR model is a flexible approach for modelling some macroeconomic time series' underlying process. The marginal likelihoods of the competing models are adequate to shed light on the model uncertainties and the existence of a deterministic time trend.

With applications of the S&P500 data sets and the U.K. /U.S. long run real exchange rates, the Gibbs Sampler algorithm is efficient[9] to provide consistent estimates for the highly parameterized dynamic GSTUR model. Considering the marginal likelihood for a high dimensional GSTUR model, Chib's method with an APF algorithm can be implemented when integrating a $T \times 1$ dimension latent variables (α_t where $t = 1$, 2, ..., T) is necessary but difficult. Therefore, this paper provide a set of tools for a complete analysis of the GSTUR model that includes MCMC estimation, diagnostics, model marginal likelihood evaluation, and estimations of the latent data α_t.

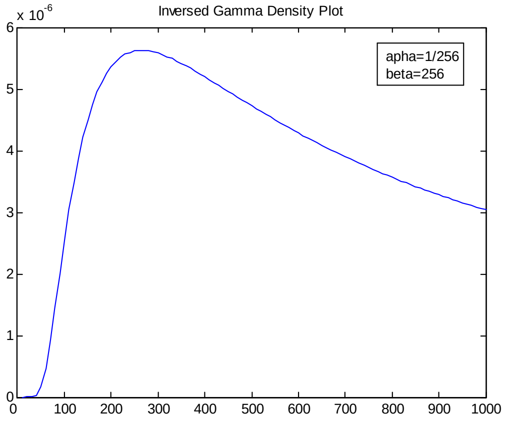
An analysis of the S&P500 stock prices series suggests that the persistence has shifted for several times within the sample. The unrestricted GSTUR model is the most favoured model, which indicates a support of the deterministic time trend. Therefore, excluding the possibility of a deterministic trend may mislead the inferences. We propose that the underlying process of the S&P500 series should be modelled with a more realistic kind, such as a combination of a drift, a deterministic time trend, and a time varying persistence with roots varying stochastically. A simple analysis of the monthly U.K.

[9] The 'efficient' here refers to quick convergence in the MCMC algorithm, small serial correlations between the sample draws and fast movements in the sample draws.

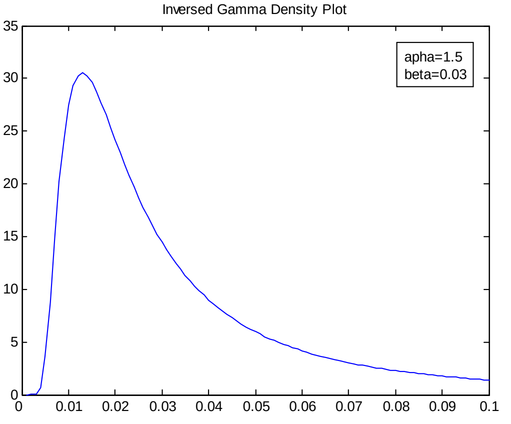
/U.S. long run real exchange rates suggests that a GSTUR model may help to resolve the puzzles relating to the puzzle of PPP. The estimated time varying stochastic roots of the series suggest that important monetary events may cause shifts in the persistence of the real exchange rates.

APPENDIX A: PRIOR ELICITATIONS

To specify an appropriate prior distribution that adequately reflects available prior information on the parameters is important. As the form of prior $p\left(\sigma_\varepsilon^2\right)$ is chosen as an Inversed Gamma distribution $f_\Gamma^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ and the form of prior $p\left(\sigma_\eta^2\right)$ is chosen as $f_\Gamma^{-1}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$. How to selecte the values of $\underline{\alpha}_\varepsilon$, $\underline{\beta}_\varepsilon$, $\underline{\alpha}_\eta$ and $\underline{\beta}_\eta$ became very important. We allow the variance σ_ε^2 in the measurement equation varies with a big range, then $\underline{\alpha}_\varepsilon = 1/256$ and $\underline{\beta}_\varepsilon = 256$. As we presume the variance σ_η^2 in the transition equation varies with a small range and centered around 0.01, we choose $\underline{\alpha}_\eta = 1.5$ and $\underline{\beta}_\eta = 0.03$. The following graphs plot the Inversed Gamma distributions $f_\Gamma^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ and $f_\Gamma^{-1}\left(\underline{\alpha}_\eta, \underline{\beta}_\eta\right)$ with above selected values.



Prior Elicitation for $f_\Gamma^{-1}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$



Prior Elicitations for $f_{\Gamma}^{-1}\left(\underline{\alpha}_{\eta}, \underline{\beta}_{\eta}\right)$

APPENDIX B: FUNCTIONS RELATED TO CONDITIONAL DRAWS FOR α

REFERENCES

[1] Busetti, F and A. M. R. Taylor. (2004). Tests of Stationarity Against a Change in Persistence. Journal of Econometrics 123, 33-66.

[2] Caner, M and B. E. Hansen. (2001). Threshold Autoregression with a Unit Root. Econometrica 69, 1555-1596.

[3] Carlin, B P. and S, Chib. (1995). Bayesian Model Choice Via Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society.Series B (Methodological) 57, 473-484.

[4] Chib, S. (1995). Marginal Likelihood from the Gibbs Output. Journal of the American Statistical Association 90, 1313-1321.

[5] Devroye, L. (1986). Non-Uniform Random Variants Generation.

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

[7] Gelfand, A. E. and A. F. M. Smith. (1990). Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85, 398-409.

[8] Geweke, J (1989). Bayesian Inference in Econometric Models using Monte Carlo Integration. Econometrica 57, 1317-1339.

[9] Granger, C. W. J. and N. R. Swanson. (1997). An Introduction to Stochastic Unit-Root Processes. Journal of Econometrics 80, 35-62.

[10] Han. C and B. P. Carlin. (2001). Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review. Journal of the American Statistical Association 96, 1122-1132.

[11] Harvey, D I.; S. J. Leybourne and A. M. R. Taylor. (2006). Modified Tests for a Change in Persistence. Journal of Econometrics 134, 441-469.

[12] Jones, C. R. and M. J. Marriott. (1999). A Bayesian Analysis of Stochastic Unit Root Models. Bayesian Statistics 6, 785-794.

[13] Judge, G. G. (1985). The Theory and Practice of Econometrics.

[14] Kapetanios, G; Y. Shin and A. Snell. (2003). Testing for a Unit Root in the Nonlinear STAR Framework. Journal of Econometrics 112, 359-379.

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

[16] Kim, J.Y. (2000). Detection of Change in Persistence of a Linear Time Series. Journal of Econometrics 95, 97-116.

[17] –, and R. B. Amador. (2002/8). Corrigendum to "Detection of Change in Persistence of a Linear Time Series". Journal of Econometrics 109, 389-392.

[18] Koop, G. (1994/7). An Objective Bayesian Analysis of Common Stochastic Trends in International Stock Prices and Exchange Rates. Journal of Empirical Finance 1, 343-364.

[19] –, E. Ley; J. Osiewalski and M. F. J. Steel. (1997/0). Bayesian Analysis of Long Memory and Persistence using ARFIMA Models. Journal of Econometrics 76, 149-169.

[20] –, and S. M. Potter. (1999/2). Bayes Factors and Nonlinearity: Evidence from Economic Time Series. Journal of Econometrics 88, 251-281.

[21] –, 2004. Bayesian Econometrics, Wiely

[22] Kwiatkowski, D.; P. C. B. Phillips; P. Schmidt and Y. Shin. (1992). Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root. Journal of Econometrics 54, 159-178.

[23] L.A.Gil-Alana and P.M.Robinson. (1997).Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series. Journal of Econometrics 80, 241-268.

[24] Nelson, C. R. and C. I. Plosser. (1982). Trends and Random Walks in Macroeconomic Time Series Journal of Monetary Economics 10, 139-162.

[25] Newbold, P.; S. Leybourne and M. E. Wohar. (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, 85-102.

[26] Pitt, M. K. and N. Shephard. (1999). Filtering Via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association 94, 590-599.

[27] Poirier, D. J. (1995). Intermediate Statistics and Econometrics: A Comparative Approach.

[28] Tanner, M. A. and W. H., Wong. (1987). The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association 82, 528-540.

[29] Taylor, M. P.; D. A. Peel and L. Sarno. (2001). Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles. International Economic Review 42, 1015-1042.

[30] van Dijk, D.; Tervirta T. and P. H. Franses. (2002). Smooth Transition Autoregressive Models — A Survey of Recent Developments. Econometric Reviews 21, 1-47.

values for $\vartheta(\phi)$ $p=1$	
t	$\vartheta(\phi)$
$t \in [p+1, T-l+p-1]$	$1+\phi_1^2$
$t = T-l+p$	1

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

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

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

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

TABLE 7
Functions for Sampling apha