

Yang, Fuyu; Leon-Gonzalez, Roberto

**Working Paper**

## Bayesian estimation and model selection in the generalised stochastic unit root model

SFB 649 Discussion Paper, No. 2010-006

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Yang, Fuyu; Leon-Gonzalez, Roberto (2010) : Bayesian estimation and model selection in the generalised stochastic unit root model, SFB 649 Discussion Paper, No. 2010-006, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/39321>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

SFB 649 Discussion Paper 2010-006

# Bayesian Estimation and Model Selection in the Generalised Stochastic Unit Root Model

Fuyu Yang\*  
Roberto Leon-Gonzalez\*\*



\* Humboldt-Universität zu Berlin, Germany

\*\*National Graduate Institute for Policy Studies Tokyo, Japan

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

# Bayesian Estimation and Model Selection in the Generalised Stochastic Unit Root Model

Fuyu Yang\*

Institute for Statistics and Econometrics  
Humboldt-Universität zu Berlin

Roberto Leon-Gonzalez<sup>†</sup>

National Graduate Institute for Policy Studies

January, 2010

## Abstract

We develop Bayesian techniques for estimation and model comparison in a novel Generalised Stochastic Unit Root (GSTUR) model. This allows us to investigate the presence of a deterministic time trend in economic series, while allowing the degree of persistence to change over time. In particular the model allows for shifts from stationarity  $I(0)$  to nonstationarity  $I(1)$  or vice versa. The empirical analysis demonstrates that the GSTUR model provides new insights on the properties of some macroeconomic time series such as stock market indices, inflation and exchange rates.

*Keywords:* Stochastic Unit Root, MCMC, Bayesian

*JEL classification:* C11, C32

## Acknowledgements

Leon-Gonzalez would like to thank the Leverhulme Trust for financial support under Grant F/00 273/J. Fuyu Yang would like to thank the financial support from the EU Commission through MRTN-CT-2006-034270 COMISEF as well as the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

---

\*Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin. Email: fuyu.yang@wiwi.hu-berlin.de. Address: Spandauer Str. 1, D-10099 Berlin, Germany.

<sup>†</sup>Fellow of the Rimini Centre for Economic Analysis.

# 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 that imply the stationarity of some series (e.g. Purchasing Power Parity). 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 empirical evidence to economic theory. This paper develops Bayesian techniques to test for a deterministic time trend while allowing changes in the degree of persistence over time. For this purpose we use a Generalised Stochastic Unit Root (GSTUR) model:

$$\nu_t = y_t - \delta t - \gamma \quad (1.1)$$

$$\nu_t = \exp(\alpha_t)\nu_{t-1} + \sum_{i=1}^l \lambda_i \Delta \nu_{t-i} + \varepsilon_t \quad (1.2)$$

$$\alpha_t = \phi_0 + \phi_1\alpha_{t-1} + \dots + \phi_p\alpha_{t-p} + \eta_t \quad (1.3)$$

where  $\varepsilon_t$  is  $i.i.d.N(0, \sigma_\varepsilon^2)$  and  $\eta_t$  is  $i.i.d.N(0, \sigma_\eta^2)$ . This is a generalization of the STUR model proposed by Granger and Swanson (1997):

$$y_t = \exp(\alpha_t)y_{t-1} + \varepsilon_t \quad (1.4)$$

$$\alpha_t = \phi_0 + \phi_1\alpha_{t-1} + \eta_t \quad (1.5)$$

where  $\varepsilon_t$  is  $i.i.d.N(0, \sigma_\varepsilon^2)$  and  $\eta_t$  is  $i.i.d.N(0, \sigma_\eta^2)$ .

One main distinctive feature of the STUR model is that it allows for the persistence of macroeconomic series to vary with time. This time-varying persistence could be a characteristic of series that appear to be nonstationary after differencing or detrending. There is empirical evidence that persistence changes over time in some U.S. macroeconomic time series (Kim 2000, Kim 2002, Busetti and Taylor 2004, Harvey et al. 2006). Thus, tests for a deterministic time trend hypothesis that ignore changes in persistence might lead to wrong conclusions. Therefore, previous evidence as to whether a macroeconomic series is trend stationary (TS) or difference stationary (DS) or neither (see Newbold et al. 2001) needs to be reconsidered. The crucial questions are: 1. how sure are we that economic time series have a deterministic trend when the persistence is time-varying and 2. whether the variations of persistence correspond with historical events. As the STUR model, the GSTUR model allows for a time-varying degree of persistence. This allows the process to be I(1) at some periods of time and I(0) at others. Furthermore, the GSTUR generalizes the STUR model by allowing for a deterministic trend and a more complex lag structure in the measurement equation and transition equation.

While modelling the changes in persistence as a stochastic process seems attractive, the estimation involved presents computational challenges. One motivation of using Bayesian techniques is that estimations for this highly parameterized model can be achieved by Markov Chain Monte Carlo (MCMC) techniques. Granger and Swanson (1997) used two methods to estimate the parameters in a STUR model (equations 1.4 and 1.5), which produced ‘wild estimates’ that are ‘fairly imprecise’. Simulations via MCMC techniques, after

passing some diagnostic tests (e.g. Carlin and Chib 1995, Geweke 1989), can shed light on the posterior distribution properties for any feature of interest. Jones and Marriott (1999) provided a Bayesian method for parameter estimations for the STUR model. In this paper, Bayesian techniques for estimation and model comparison in a GSTUR model are developed.

The remainder of the paper is organized as follows. Section 2 describes the methods for Bayesian estimation and model selection. Section 3 presents the empirical results using Nelson and Plosser's S&P 500 series, the U.K. /U.S. long run exchange rates and UK inflation rates. Section 4 briefly concludes.

## 2 Bayesian Inference

Equation 3 defines the time-varying persistence parameter  $\alpha_t$  as an  $AR(p)$  process, which we assume to be stationary. Thus the unconditional mean  $\mu_\alpha$  of the stationary process  $\alpha_t$  is:

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

Let

$$\rho_t = \exp(\alpha_t) \tag{2.1}$$

We begin by introducing some notation:  $F_t$  denotes the history of  $y_t$  up to time  $t$ ,  $F_t = (y_1, \dots, y_t)'$ ,  $y$  denotes the whole sample of observations with a sample size of  $N$ ,  $y = (y_1, y_2, \dots, y_n)'$ . Let us also define  $\alpha = (\alpha_{1-p}, \dots, \alpha_0, \alpha_1, \dots, \alpha_{T-1}, \alpha_T)'$  as the vector containing all unobserved stochastic roots over the time  $T$  period. Similarly define the vector  $\rho = (\rho_{1-p}, \dots, \rho_0, \rho_1, \dots, \rho_{T-1}, \rho_T)'$  associated with  $\alpha$ . Define also  $\alpha_{initial} = (\alpha_{1-p}, \dots, \alpha_0)'$ ,  $\phi = (\phi_1, \dots, \phi_p)'$  and  $\lambda = (\lambda_1, \dots, \lambda_l)'$ . The error precisions are denoted as  $h_\varepsilon = \sigma_\varepsilon^{-2}$  and  $h_\eta = \sigma_\eta^{-2}$ .  $\theta = (\gamma, \delta, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2)$  stands for the vector containing all the parameters of interest. Note that  $\rho_t$  determines the degree of persistence and varies stochastically in the GSTUR process.

Note that the process is stationary (I(0)) if  $\rho_t < 1$  and not stationary if  $\rho_t \geq 1$ . Note also that the random walk (RW) model is nested within the GSTUR model at the point where  $\mu_\alpha = 0$  and  $\sigma_\eta^2 = 0$ , such that  $\rho_t$  will be always equal to 1. Hence,  $(\mu_\alpha, \rho_t, \delta)$  are parameters of special interest.

### 2.1 Bayesian Model Estimation

With few exceptions detailed below, we adopt the prior specification proposed in Jones and Marriott (1999) (JM hereafter):  $\mu_\alpha \sim N(\underline{\mu}_\alpha, \underline{V}_\alpha)$ ,  $\{\gamma, \delta\} \sim N(\underline{\mu}_\Omega, \underline{V}_\Omega)$  and  $\lambda \sim N(\underline{\mu}_\lambda, \underline{V}_\lambda)$ . Since  $\alpha_t$  is restricted to be a stationary  $AR(p)$  process, the inverse characteristic roots of the polynomial:

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

should all lie outside of the unit circle. Let  $1(\|z^*\| > 1)$  be an indicator function for the event that  $\phi_i : i = 1, \dots, p$  jointly satisfy the stationarity condition. The prior density of  $\phi$  is  $p(\phi) = (\underline{C})^{-1} f_N(\underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}) 1(\|z^*\| > 1)$ , where  $f_N(\cdot)$  is a multivariate normal density

and  $\underline{C}$  is the normalizing constant. The prior densities for error precisions are chosen as  $p(h_\varepsilon) = f_{\Gamma\varepsilon}(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon)$  and  $p(h_\eta) = f_{\Gamma\eta}(\underline{\alpha}_\eta, \underline{\beta}_\eta)$ , where  $f_\Gamma(\cdot)$  denotes the density of the Gamma distribution.

Table (1) shows our choice for prior parameters in the empirical application. The parameter  $\mu_\alpha$  varies around  $\ln 0.9$  with a small variance  $\underline{V}_\alpha$  and is independent of  $\sigma_\eta^2$  (unlike in JM, where  $\mu_\alpha | \sigma_\eta^2 \sim f_N(0, \sigma_\eta^2)$ ). Besides, JM selected ‘diffuse’ priors for  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$  that merely indicate that  $0 < \sigma_\varepsilon^2 < \infty$  and  $0 < \sigma_\eta^2 < \infty$ . Instead, we want a prior that captures the prior belief that  $\rho_t$  varies slowly over time. Accordingly, the prior for  $\sigma_\eta^2$  puts most of the probability mass on small values. We fix the prior parameters of  $\sigma_\varepsilon^2$  so that the prior probability of extremely large values is small.

Table 1: Summarized Prior Properties in GSTUR

parameters	Selected Values in the Prior			
$\mu_\alpha$	$\underline{\mu}_\alpha$	$\ln 0.9$	$\underline{V}_\alpha$	$0.1^2$
$\phi_i$	$\underline{\mu}_{\phi_i}$	1	$\underline{V}_{\phi_i}$	0.1
$h_\varepsilon$	$\underline{\alpha}_\varepsilon$	1.1	$\underline{\beta}_\varepsilon$	0.2
$h_\eta$	$\underline{\alpha}_\eta$	1.5	$\underline{\beta}_\eta$	2.5
$\Omega$	$\underline{\mu}_\Omega$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\underline{V}_\Omega$	$10^4 \text{eye}(2)$
$\lambda$	$\underline{\mu}_\lambda$	$(0, \dots, 0)'$	$\underline{V}_\lambda$	$10^4 \text{eye}(i)$

Given the above prior information, we are able to derive the conditional posterior distributions that can be used in the MCMC algorithm. To facilitate computations, we will work with the augmented likelihood, such that the posterior for  $\theta$  can be written as:

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

All conditional posterior densities have standard forms and can be sampled directly, except for that corresponding to  $\alpha_t$ , which is non-standard (see Appendix for the detail form of this density). JM applied a ratio of uniforms method (see Devroye, 1986) to sample  $\alpha_t$ . As the posterior conditional of  $\alpha_t$  is similar to a normal density, the Metropolis-Hastings (M-H) algorithm can be implemented to draw values of  $\alpha_t$ . As shown in the empirical section, the proposed Metropolis-within-Gibbs algorithm exhibits fast convergence. The following MCMC procedure can be used to simulate from the posterior distribution in the GSTUR model. The exact expressions for the parameters of the conditional posteriors can be found in the appendix.

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

1. Give initial values for  $\sigma_\varepsilon^2, \sigma_\eta^2, \mu_\alpha, \phi, \alpha, \lambda, \Omega$
2. Repeat the steps  $a - g$  during  $S$  iterations.
  - (a) Sample  $\sigma_\varepsilon^2$  from  $(\sigma_\varepsilon^2 | y, \Omega, \lambda, \alpha, \phi, \mu_\alpha, \sigma_\eta^2)$ , which has density  $f_{\Gamma\varepsilon}^{-1}(\bar{\alpha}_\varepsilon, \bar{\beta}_\varepsilon)$
  - (b) Sample  $\sigma_\eta^2$  from  $(\sigma_\eta^2 | y, \Omega, \lambda, \alpha, \phi, \mu_\alpha, \sigma_\varepsilon^2)$ , which has density  $f_{\Gamma\eta}^{-1}(\bar{\alpha}_\eta, \bar{\beta}_\eta)$
  - (c) Sample  $\mu_\alpha$  from  $(\mu_\alpha | y, \Omega, \lambda, \alpha, \phi, \sigma_\varepsilon^2, \sigma_\eta^2)$ , which has density  $f_N(\bar{\mu}_\alpha, \bar{V}_{\mu_\alpha})$

- (d) Sample  $\phi$  from  $(\phi | y, \Omega, \lambda, \alpha, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2)$ , which has density  $f_N(\bar{\mu}_\phi, \bar{V}_\phi)$  truncated to the stationary region. If  $\phi$  satisfies the stationary condition  $\|z_j\| > 1$ , continue to step e. Otherwise draw the whole vector  $\phi$  again.
- (e) Sample  $\alpha_t$  from  $p(\alpha_t | y, \Omega, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2, \alpha_{-t})$  using the independent chain M-H algorithm, in which a univariate  $t$ -density is chosen as the candidate generating density. Note that we use the notation  $\alpha_{-t} = (\alpha_{1-p}, \dots, \alpha_{t-1}, \alpha_{t+1}, \dots, \alpha_{T-1}, \alpha_T)'$
- (f) Sample  $\lambda$  from  $\lambda | y, \Omega, \alpha, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2$ , using the density  $f_{MN}(\bar{\mu}_\lambda, \bar{V}_\lambda)$
- (g) Sample  $\Omega$  from  $\Omega | y, \alpha, \lambda, \phi, \mu_\alpha, \sigma_\varepsilon^2, \sigma_\eta^2$ , using the density  $f_{MN}(\bar{\mu}_\Omega, \bar{V}_\Omega)$

## 2.2 Bayesian Model Comparison

The most important aspects in model specification relate to the existence of a deterministic time trend (whether  $\delta = 0$ ), the time variation in  $\rho$  (whether  $\rho$  is time-invariant) and the number of lags in the measurement equation (the value of  $l$ ). The Bayesian approach provides a compelling framework to tackle model uncertainty (e.g. Kass and Raftery, 1995) and over-parameterization (e.g. Koop and Potter, 1999). According to Kass and Raftery (1995), the Bayes Factor  $B_{ij}$  that compares model  $M_i$  and model  $M_j$  is expressed as

$$B_{ij} = \frac{p(y|M_i)}{p(y|M_j)} \quad (2.2)$$

where  $p(y|M_i)$  is the marginal likelihood in model  $M_i$ . The strength of evidence in favour of model  $M_i$  versus  $M_j$  can be evaluated following the recommendations of Kass and Raftery (1995), which are summarized in Table (2):

$\log_{10} B_{ij}$	$B_{ij}$	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

As the GSTUR model is highly dimensional, the marginal likelihood of the GSTUR model cannot be obtained with a straightforward analytical integration method. In the case where a Gibbs Sampler is already implemented and all the full posterior conditional densities are known, it is possible to approximate the marginal likelihood of each competing model using the posterior draws and the approach introduced by Chib (1995).

The idea of the Chib method starts from the basic ‘marginal likelihood identity’. The marginal density of  $y = (y_2, \dots, y_n)'$  can be written as:

$$p(y) = \frac{p(\theta) \prod_{t=2}^n p(y_t | \theta)}{p(\theta | y)} \quad (2.3)$$

where the numerator is the product of the sampling density and the prior density (including the integrating constant  $\underline{C}$ ) and the denominator is the posterior density of  $\theta$ . The constant  $\underline{C}$  can be evaluated by simulating from the untruncated prior of  $\phi$ , and calculating the

proportion of times that the draws verify the stationarity condition (e.g. Judge et al. (1985, pp.128)). The conditional posterior density of  $\phi$  is also subject to truncation and so we explain how to calculate the corresponding normalizing constant. As the above ‘marginal likelihood identity’ holds for any  $\theta$ , say  $\theta^*$ , the log marginal likelihood can be approximated as:

$$\ln p(y) \propto \sum_{t=2}^n \ln p(y_t | \theta^*, F_{t-1}) + \ln [p(\theta^*)] - \ln [p(\theta^* | y)] \quad (2.4)$$

where  $p(\theta^*)$  is easily obtained by evaluating the prior densities at  $\theta^{*1}$ . For greater accuracy  $\theta^*$  should be a point of high posterior density. In this paper it is chosen to be the posterior mean. The posterior ordinates  $\ln p(\theta^* | y)$  can be obtained using a marginal/conditional decomposition together with the outputs from the original and subsequent ‘reduced MCMC runs’. Using the marginal/conditional decomposition:

$$\begin{aligned} p(\phi^*, \mu_\alpha^*, h_\eta^*, \gamma^* \delta^*, \lambda^*, h_\varepsilon^* | y) &\propto p(\mu_\alpha^* | y) p(h_\eta^* | y, \mu_\alpha^*) \\ &(\gamma^*, \delta^* | y, \mu_\alpha^*, h_\eta^*) p(\lambda^* | y, \mu_\alpha^*, h_\eta^*, \gamma^* \delta^*) \\ &p(h_\varepsilon^* | y, \mu_\alpha^*, h_\eta^*, \gamma^* \delta^*, \lambda^*) p(\phi^* | y, \mu_\alpha^*, h_\eta^*, \gamma^* \delta^*, \lambda^*, h_\varepsilon^*) \end{aligned} \quad (2.5)$$

The first term  $p(\mu_\alpha^* | y)$  in Equation (2.5) can be estimated from the MCMC algorithm outputs as

$$p(\mu_\alpha^* | y) \simeq \frac{1}{S} \sum_{s=1}^S p(\mu_\alpha^* | y, \alpha^{(s)}, \mu_\alpha^{(s)}, h_\eta^{(s)}, \Omega^{(s)}, \lambda^{(s)}, h_\varepsilon^{(s)}, \phi^{(s)}) \quad (2.6)$$

The same calculation is done to obtain  $p(h_\eta^* | y, \mu_\alpha^*)$ , but the draws come from a reduced MCMC run. The reduced MCMC run consists in running again Algorithm 1, but fixing  $\mu_\alpha$  to be equal to the posterior mean  $\mu_\alpha^*$  in every iteration. The other components on the right hand of Equation (2.5) can be approximated in an analogous manner using draws from reduced MCMC runs with the appropriate variables fixed to the corresponding values in  $\theta^*$ . The normalizing constant  $p(\phi^* | y, \alpha, \mu_\alpha^*, h_\eta^*, \gamma^*, \delta^*, \lambda^*, h_\varepsilon^*)$  can be approximated by the number of draws that verify the stationarity restriction.

As shown in appendix B, the method of auxiliary particle filtering (APF) (Pitt and Shephard (1999)) can be adapted to estimate the value of the likelihood  $\sum_{t=2}^n \ln p(y_t | \theta^*, F_{t-1})$ .

Thus, the marginal likelihood identity (2.3) allows us to estimate the log marginal likelihood of the GSTUR model with or without restrictions. We also would like to compare the simpler linear Random Walk (RW) model with the GSTUR model. The marginal likelihood of the RW model can be evaluated analytically:

$$p_{RW}(y) = \int p(y | \sigma_\varepsilon^2) p(\sigma_\varepsilon^2) d\sigma_\varepsilon^2 \quad (2.7)$$

In the RW model,  $\sigma_\varepsilon^2$  is the only parameter that has to be estimated. If  $h_\varepsilon = \sigma_\varepsilon^{-2}$  and the prior chosen for  $h_\varepsilon$  is a Gamma distribution  $h_\varepsilon \sim \Gamma(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon)$ , the marginal likelihood of the RW model would depend only on the values of  $\underline{\alpha}_\varepsilon$  and  $\underline{\beta}_\varepsilon$ . For the purpose of model comparison, it is sensible to choose the same prior in competing models for the common parameters. Thus, the values of  $\underline{\alpha}_\varepsilon$  and  $\underline{\beta}_\varepsilon$  will be chosen to be the same as those for  $\sigma_\varepsilon^{-2}$  in the GSTUR model (see Table 1).

<sup>1</sup>The exact mathematical expression for the prior densities can be found in the Appendix.



### 3 Empirical Illustrations with Generalised STUR

An application to the series of Standard & Poor 500 indices (S&P500) indicates that, once we allow for stochastic unit roots, a deterministic time trend might exist in the S&P500 series. A further application to U.K. /U.S. long-run exchange rate indicates that changes in persistence coincide with monetary events. Finally, using the UK inflation series, we demonstrate that the GSTUR captures well the inflation dynamics and we assess the out-of-sample prediction error.

#### 3.1 Empirical Results with Stock Price

We use the S&P500 annual data set, measured in logarithms, for the period 1877 to 1988. 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). This data set has also been used by Jones and Marriott (1999) to estimate the original Stochastic Unit Root model (the simplest form of the GSTUR in equation 1.4 and 1.5). Here, in addition to estimate the more general GSTUR model, we test hypotheses using posterior probabilities.

##### 3.1.1 Estimation

To ensure that the effects of the starting values in the MCMC algorithms are insignificant, we take 25,000 draws after discarding the initial 5,000. The MCMC convergence diagnostic results from the Numerical Standard Errors (NSE)<sup>2</sup> and the Convergence Diagnostic values (CD) show that the number of iterations used is more than sufficient. The correlogram (autocorrelation function) plots serial correlations of the draws from the posterior simulator. Fig (1 and 2) indicate that, for all the parameters of interest, there is no significant autocorrelation at lag lengths larger than 15. Thus, the quick decaying autocorrelation indicates quick movements in the sampled draws and fast convergence.

Table 3: Estimates: GSTUR with an Application to SP500

	<i>Prior</i>		<i>Posterior</i>						
	<i>Mean</i>	<i>StDev</i>	<i>Mean</i>	<i>StDev</i>	<i>CD</i>	<i>NSE<sub>15</sub></i>	<i>Median</i>	<i>.95HPDI</i>	
$\mu_\alpha$	ln 0.9	0.1 <sup>2</sup>	-0.1176	0.0678	-0.1168	0.0031	-0.1155	-0.2315	-0.0104
$\sigma_\eta^2$	-	-	0.0382	0.0122	0.2324	0.0002	0.0360	0.0227	0.0613
$\sigma_\varepsilon^2$	-	-	0.1160	0.0159	-0.6118	0.0001	0.1147	0.0926	0.1445
$\phi_1$	†	†	0.1079	0.1321	-1.2492	0.0008	0.1096	-0.109	0.3234
$\gamma$	0	10 <sup>6</sup>	0.7649	0.3969	-0.2616	0.0035	0.7709	0.107	1.4092
$\delta$	0	10 <sup>6</sup>	0.0345	0.0057	0.192	0	0.0346	0.0252	0.0438
$\lambda_1$	0	10 <sup>6</sup>	0.2528	0.2109	-0.1585	0.0021	0.2516	-0.0931	0.6003

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

- : see Table (1) for prior descriptions

The summary statistics of Table (3) reports the estimated results and efficiency diagnostics of the MCMC algorithm. According to the CD value and NSE values, the MCMC algorithm converges for all the parameters of interest. The posterior estimates show that

<sup>2</sup>The number of lags used to calculate NSE is 15% of the size of the sample.

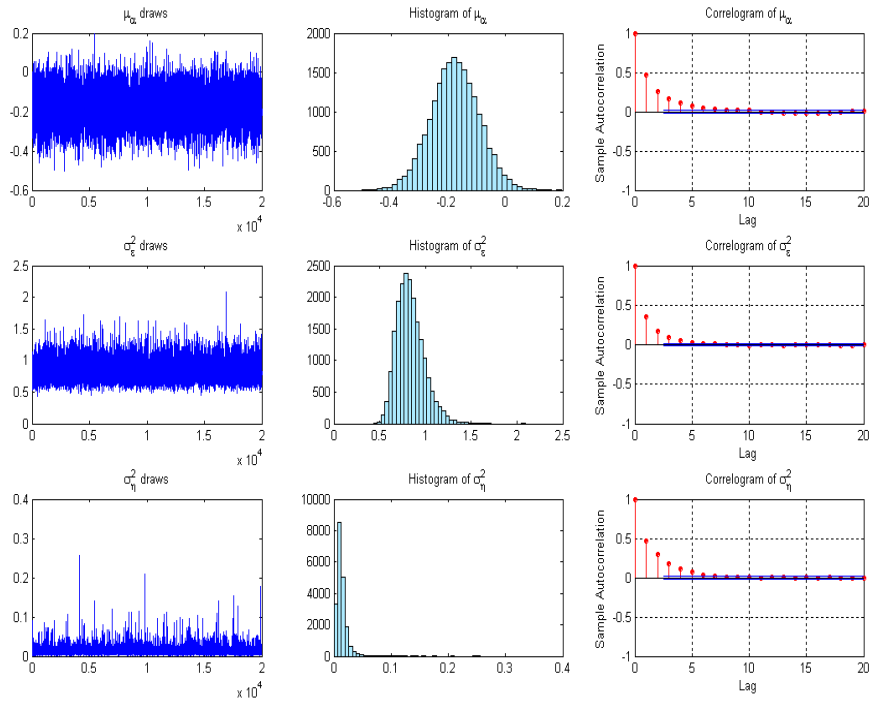


Figure 1: SP500 with GSTUR: Posterior Draws of  $\mu_\alpha$ ,  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$

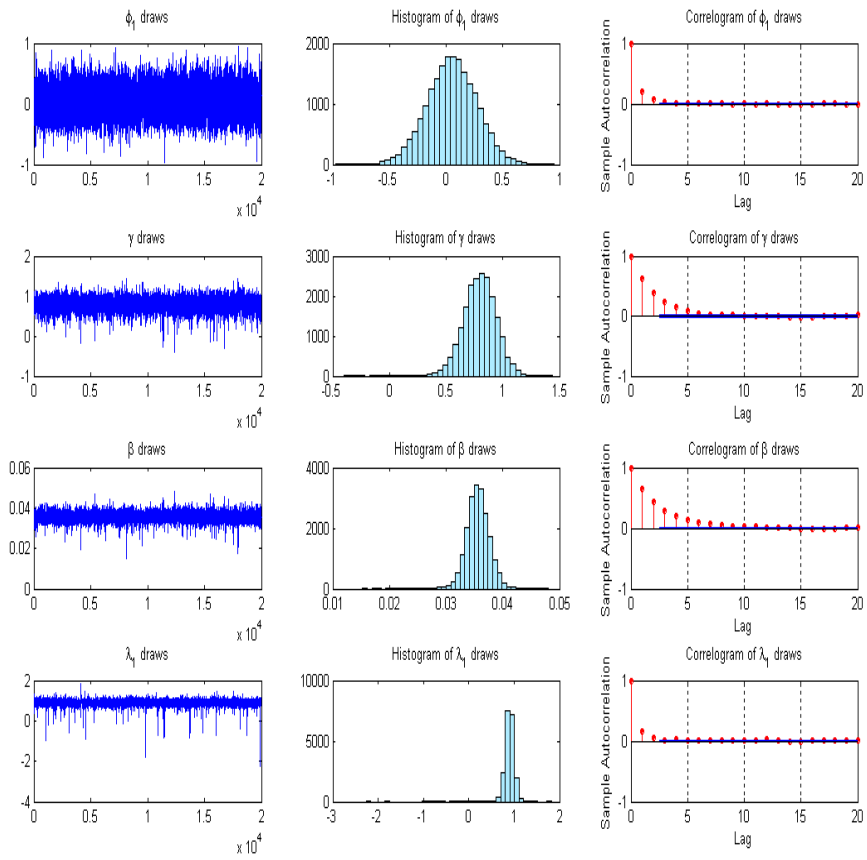


Figure 2: SP500 with GSTUR: Posterior Draws  $\phi_1, \gamma, \beta$ , and  $\lambda_1$

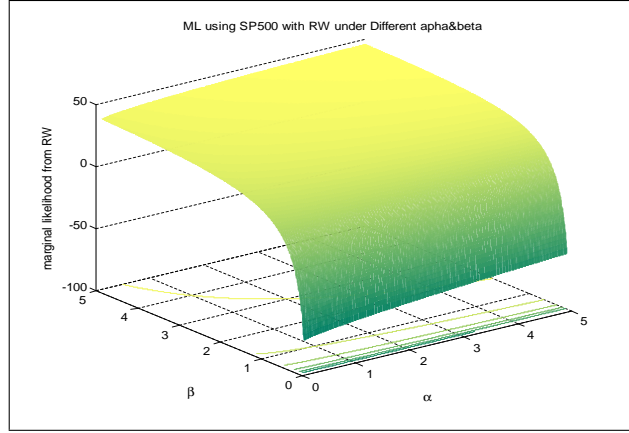


Figure 3: log Marginal Likelihood of RW with an Application of SP500

parameters are substantially different from zero. A negative  $\mu_\alpha$  and small  $\sigma_\eta^2$  suggests that the S&P500 series could be a process with Stochastic Unit Roots.

### 3.1.2 Model Selection

Imposing  $\delta = 0$  or  $\gamma = \delta = 0$  affects the estimates of stochastic roots significantly. As pointed out by Koop (1994), restricting the deterministic time trend to be zero forces any trend behavior to manifest itself stochastically, biasing the tests in favour of stochastic nonstationarity. Considering the problem of over-parameterization, it is also important to decide which parameters should be included in a good fitting model. Table (4) presents the estimated log marginal likelihood values. For a given value of  $l$  (the lag length in the measurement equation (1.2)), these values do not vary much with  $p$  (the lag length in the transition equation (1.3)). In contrast, log marginal likelihood values vary substantially with  $l$  for a given  $p$ . Thus, the lag length in the measurement equation is more important than the lag length in the transition equation.

To compare the highly parameterized nonlinear GSTUR model with the simple linear RW model, we calculate the marginal likelihood for the RW model. Because the marginal likelihood from the RW model depends on the values of  $\underline{\alpha}_\varepsilon$  and  $\underline{\beta}_\varepsilon$  in the prior ( $\sigma_\varepsilon^{-2} \sim \Gamma_\varepsilon(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon)$ ), 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), 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  $\underline{\beta}_\varepsilon = 0.2$ , which are the same as those in the prior of  $\sigma_\varepsilon^{-2} \sim f_{\Gamma_\varepsilon}(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon)$  in the GSTUR model, the log marginal likelihood is  $-34.413$ . Given that, as argued above, we choose the same prior for the common parameter  $\sigma_\varepsilon^{-2}$ , the Bayes Factors between the RW and the GSTUR model (with  $l = 1$ ,  $\delta \neq 0$ ) with a deterministic time trend ( $\delta \neq 0$ ) can be calculated as:

$$BF_{RW:GSTUR} = \frac{p(M_{RW})}{p(M_{GSTUR})} = \frac{\exp(-34.413)}{\exp(-0.2384)} = 1.4393 \times 10^{-15}$$

According to Table (2), there is very strong evidence in favour of the GSTUR model when

compared to the RW model. We may conclude that the sample series has a 99.9% probability of being a stochastic unit root process.

To visualize the changes in persistence over the sample period ( $t = 1878 - 1988$ ), the estimated time-varying roots  $\rho_t$  from the GSTUR model are plotted below the S&P 500 data. The estimates of the roots  $\rho_t$  vary under different specifications of the constant  $\gamma$  and the deterministic time trend  $\delta$ . Fig (4) plots the estimated stochastic roots for the GSTUR ( $p = 1, l = 1$ ) model with deterministic trend ( $\delta \neq 0$ ) but no constant ( $\gamma = 0$ ). Fig (5) plots the same for the unrestricted GSTUR ( $p = 1, l = 1$ ) model ( $\delta \neq 0, \gamma \neq 0$ ).

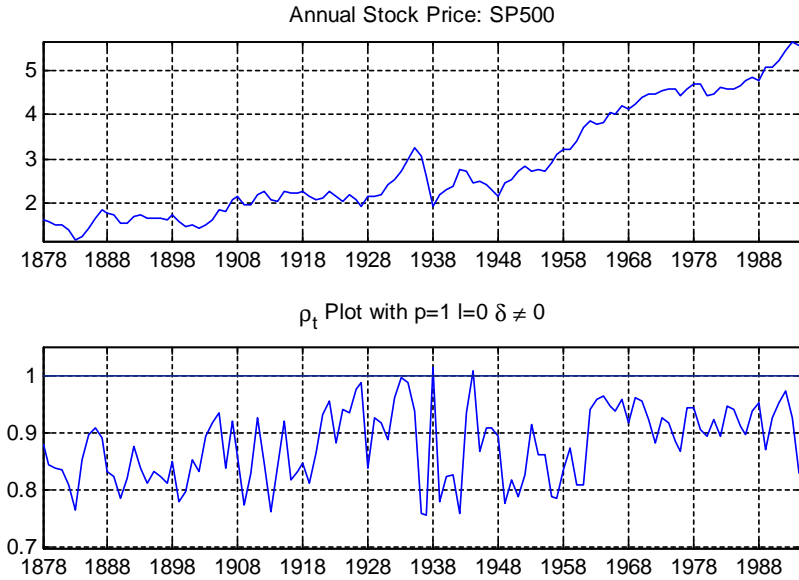


Figure 4: GSTUR with  $\delta \neq 0$  : Time-Varying Roots of SP500 1878-1988

In Figures (4 and 5) we observe that the roots go above unity occasionally. That is, 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, exhibiting an explosive behaviour. This may explain why Nelson and Plosser (1982), Kwiatkowski et al. (1992) all find significant evidence in favor of a Unit Root in the S&P 500 series.

### 3.2 Empirical Results with Long-run Real Exchange Rate

Another empirical application for the GSTUR model is to analyze the real exchange rates. We use the monthly U.K./U.S. real exchange rates from January 1885 to February 1995, with a sample size of 1,322 over a 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 at the 5 percent level.

In this section, a restricted GSTUR model ( $\gamma = \delta = 0, p = 1$  and  $l = 1$ ) is used. The algorithm uses 25,000 iterations with the first 5,000 discarded. The efficiency of the

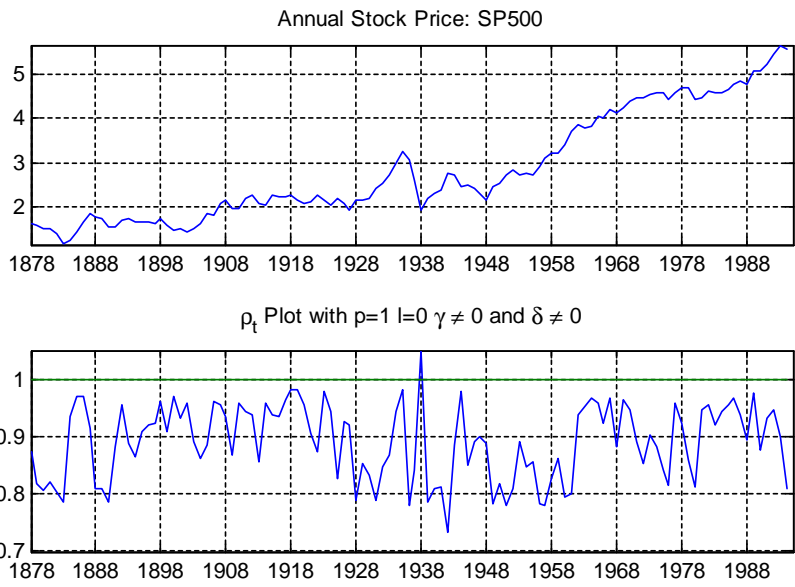


Figure 5: GSTUR with  $\gamma \neq 0$   $\delta \neq 0$  : Time-Varying Roots of SP500 1878-1988

algorithm can be assessed with the CD and NSE values. Table (5) reports the estimated results and diagnostics of the MCMC.

Figure (6) plots the U.K./U.S. real exchange rates, nominal exchange rates and estimated roots over the 111 year span. 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. However, at certain time points the roots jump to or above one. These points are marked with  $\dagger$ , and indicate nonstationary and/or explosive behaviors. Engel and Kim (1999) provide a description of the historical monetary events within the 111-year span and Figure 6 marks these events with vertical bars. Note that there is a correlation between the historical events and the  $\dagger$  marks.

### 3.3 Empirical Results with UK inflation

The third empirical application with the GSTUR model is to analyze the quarterly U.K. inflation series. Figure (7) 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. The inflation rates were controlled to vary within a small range after the introduction of 'inflation targeting' in the 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.

Figure (8) 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  $\dagger$  sign on Figure (8) marks the points at which the roots jump above one. The pattern of the stochastic roots before 1993 looks very different from that after 1993. Before 1993 the

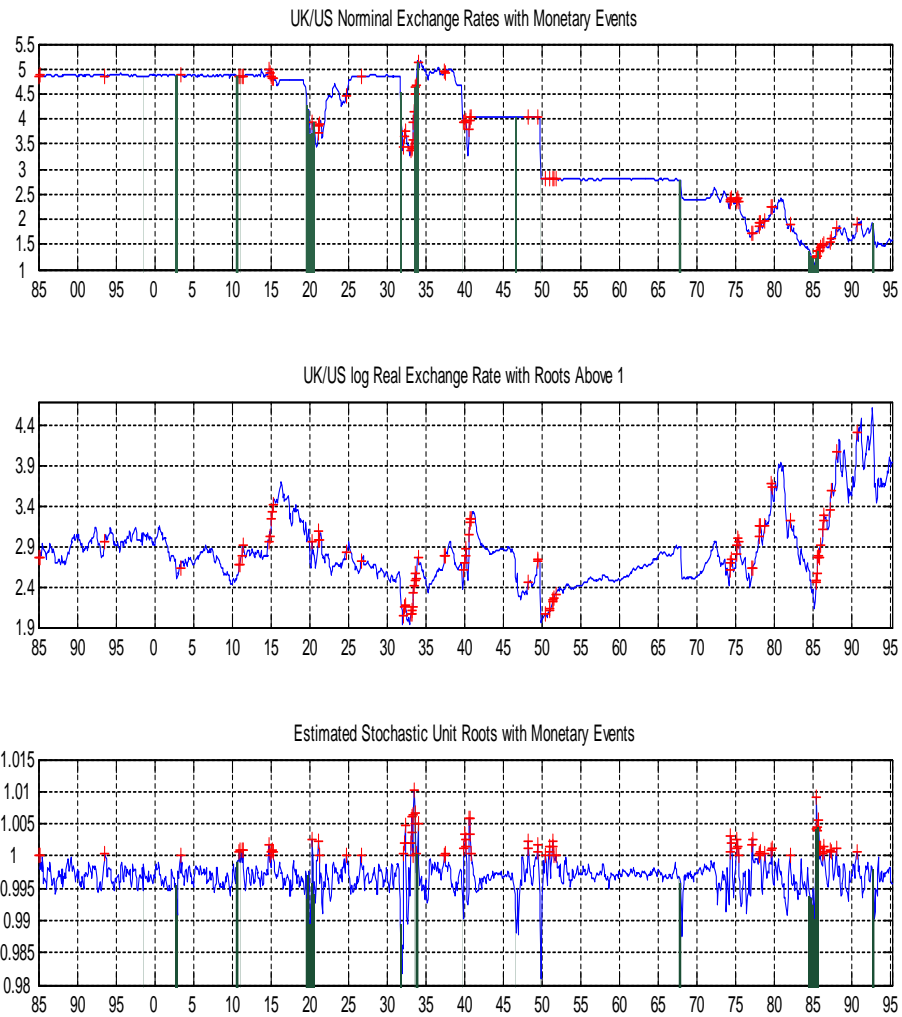


Figure 6: U.K./U.S. Long Run Exchange Rates and Estimated Stochastic Unit Roots

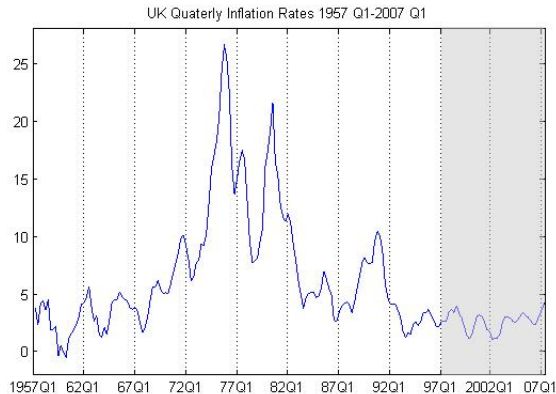


Figure 7: Plot of UK inflation series: 1957 Q1- 2007 Q1

stochastic roots jump above 1 frequently. However, the roots stayed below unity after 1993. The result from estimating a GSTUR model with UK inflation data provides good evidence of a changing persistence in the inflation underlying process, which is consistent with the findings of 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.

In the out-of-sample forecasting exercise we will compare the GSTUR model with linear models (i.e. AR and RW models). We will obtain one-step-ahead forecasts for the period 2004 Q4 to 2007 Q1. In the first exercise we will calibrate the models using observations from 1957Q1. In the second we will use only observations after the independence of the Bank of England, which is the period when inflation dynamics are simpler and could potentially be captured by a linear stationary model (e.g. AR model).

We also construct density forecast for each model. The density forecasts are represented graphically as a set of prediction intervals covering 10, 20,..., 90 percent of the probability distribution, with lighter shades for the outer bands. They are constructed in such a way that the boundaries of the intervals are the 5th, 10th,...,95th percentiles. Note that intervals are not always symmetric around the mode, which is represented with the darkest blue. With multiple-step ahead forecasts, the forecast intervals “fan out” as the forecast horizon increases. Since we focus only on one-step ahead forecasts, there are no fan out effects.

### 3.3.1 Full sample

We present the descriptive statistics of the forecast distributions in tables, which include the mean, mode, median, variance and 95% percentile of the forecast distributions. The marginal likelihood of each model is calculated conditional on the observations before the forecast. With lag length fixed at 7, Tables 6, 7,8 and 9 show the forecasts and marginal likelihood values for the GSTUR model subject to alternative restrictions on the constant and trend. Table (10) shows the same for the RW model.

Table (11) presents the MSFE for all entertained models with different specifications. The GSTUR model with lag length 7 and a constant with no trend receives the highest marginal likelihood and provides the best point forecast in terms of MSFE. It is significant



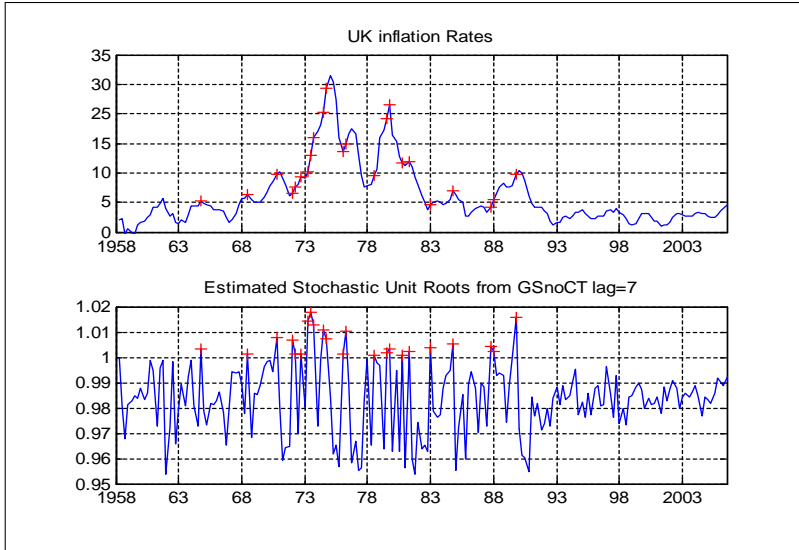


Figure 8: Simulated Roots Using the GSTUR model with No Constant or Trend with lag length at 7

that almost all specifications for the GSTUR model have substantially lower MFSFE than the RW model.

Figures 9.a, 9.b and 9.c show the density forecast plots for 3 models<sup>3</sup>: GSTUR with constant and lag length of 7, GSTUR with trend and lag length of 7, and RW model. From Figure 9.a and Figure 9.b, a GSTUR model with trend generates larger forecast variances than a model with a constant. From the forecast results, the GSTUR with a constant and trend does not improve forecast accuracy over two of the restricted GSTUR models (i.e. without constant or without trend). Moreover, the forecast variances from the GSTUR models with constant and trend are 3 times larger than those from the GSTUR model with only a constant. Thus, forecast variance and Mean Square Forecast Error (MSFE) values point in the same direction as the marginal likelihood values. Therefore, the posterior model probability of the unrestricted GSTUR model will be low and it might be better to exclude it from the forecasting averaging in this case.

<sup>3</sup>Density forecasts for other models are in the Appendix.

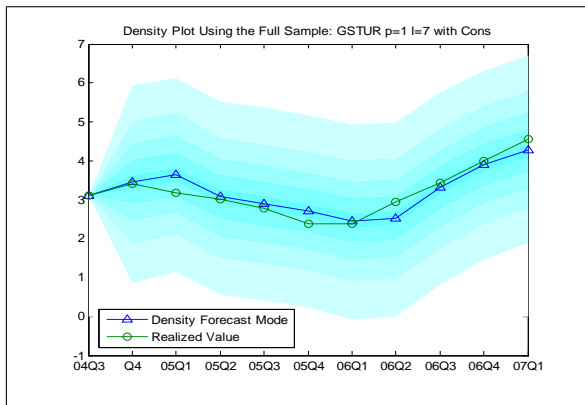


Fig 9.a GSTUR  $p=1$   $l=7$  with Cons

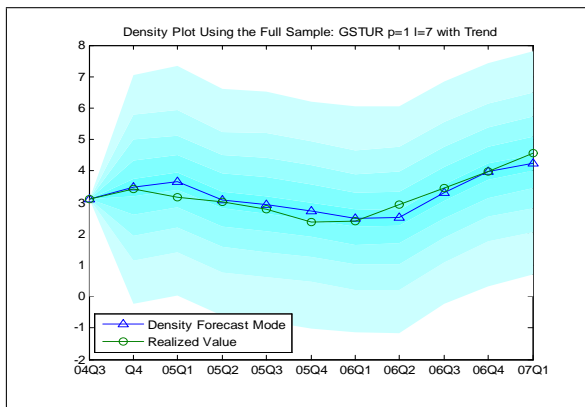


Fig 9.b GSTUR  $p=1$   $l=7$  with Trend

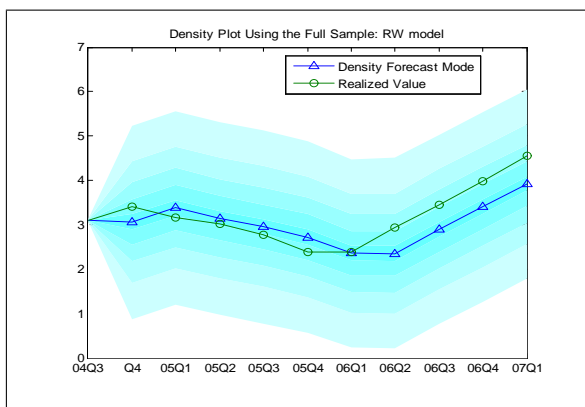


Fig 9.c RW model

### 3.3.2 Small sample

We carry out exactly the same exercise as in Section 3.3.1 but using only observations from 1999Q4. For example, to forecast inflation in 2004 Q4 we estimate the models using the period (1999 Q4 - 2004 Q3), which is 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.

Tables (12) and (13) provide descriptive statistics of the forecast distributions using the GSTUR without a constant or trend at lag 4, and the RW model. Figures 10.a, and 10.b plot the forecast densities for the GSTUR without constant or trend at lag 5, and the RW model.

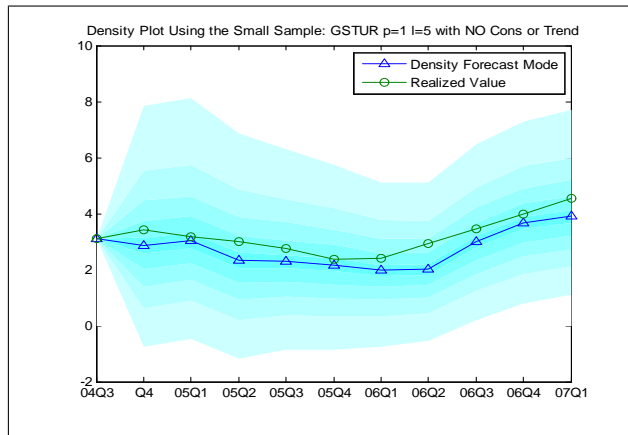


Figure 10.a: Small Sample with GSTUR no CT, lag=5

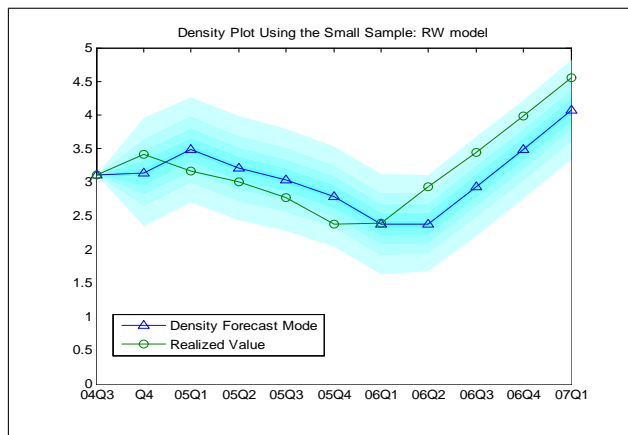


Figure 10.b: Small sample RW

According to the MSFE value in Table (14), a GSTUR without a constant or trend and 4 lags provides the best point forecasts. Thus, although the data series exhibit simple dynamics that could potentially be well captured by a linear model, the nonlinear GSTUR model still provides better point forecasts than a RW model or a stationary AR(5) model.

This indicates that the GSTUR model could capture the underlying dynamics of inflation not only when the data exhibits nonlinear dynamics, but also when the data dynamics are linear. However, using the small sample as opposed to the full sample does not improve on the point forecast accuracy of the GSTUR model (see Tables (11) and (14)), but it does improve the performance of the RW model.

If we look into Figure 10.a and Table (12), 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 slightly, the forecasting variances decrease. Thus, the big forecast uncertainty may be due to the small size of the sample. Compared to results in Table (9), where the same model specification was applied to a larger sample, the forecast variances are much larger.

Therefore, the GSTUR outperforms the RW model in both the full sample and short sample case. In the case of forecasting using the GSTUR model, it seems better to estimate the model using a large sample even if it contains a structural break. That is, using a small but more homogeneous sample might result in big forecasting uncertainties and errors.

## 4 Conclusions

The GSTUR is a flexible nonlinear model which can successfully capture the properties of some macroeconomic time series. It allows for changes in persistence and therefore can provide a better understanding on the sources of macroeconomic fluctuations as well as help to identify structural breaks. In this paper we have developed an MCMC algorithm for Bayesian parameter estimation and model comparison<sup>4</sup> and found that these computational methods worked satisfactorily when applied to real data. In particular the methods can be used to assess the evidence in favour of a deterministic trend, which is important for forecasting purposes.

This paper revisits the dispute concerning the existence of a 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 over time. Among the competing models, the GSTUR with a deterministic time trend has the highest marginal likelihood, which indicates a support for 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.

A simple analysis of the monthly U.K./U.S. long run real exchange rates over the 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.

Lastly, using UK inflation data the GSTUR class models are better in providing out-of-sample forecasts than the RW model. Using different sample sizes and periods to calibrate the model, the GSTUR model outperforms the RW model in terms of out-of-sample forecasting accuracy. Hence, the nonlinear GSTUR models are not only resilient to 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

---

<sup>4</sup>The Matlab code for estimation, model comparison, MCMC diagnosis, likelihood evaluation and estimation of  $\alpha$  is available from the authors upon request.

a GSTUR process as it may be able to accommodate breaks in persistence and provide accurate forecasts.

## A Prior and Posterior Densities

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

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

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

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

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

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

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

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

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

For the values of  $(\underline{\mu}_\alpha, \underline{V}_\alpha, \underline{\mu}_\lambda, \underline{V}_\lambda, \underline{\mu}_\Omega, \underline{V}_\Omega, \underline{\mu}_{\phi_i}, \underline{V}_{\phi_i}, \underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon, \underline{\alpha}_\eta, \underline{\beta}_\eta)'$ , refer to Table (1).

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

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

Since

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

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

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

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

and

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

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

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

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

$$\begin{aligned}
p(\theta, \alpha | y) &\propto \prod_{t=2}^n p(y_t | \alpha, \theta, F_{t-1}) p(\alpha, \theta) \\
&\propto \prod_{t=2}^n p(\alpha_t | \alpha_{-t}, \theta) p(\theta) \prod_{t=2}^n p(y_t | \alpha, \theta, F_{t-1}) \\
&\quad \cdot \frac{1}{(2\pi\sigma_\varepsilon^2)^{\frac{n-1}{2}}} \exp \left\{ \left( \begin{array}{c} -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^n \\ y_t - \exp(\alpha_t)y_{t-1} - \sum_{i=1}^l \lambda_i \Delta y_{t-i} - \\ \gamma [1 - \exp(\alpha_t)] - \\ \delta \left[ t - \exp(\alpha_t)t + \exp(\alpha_t) - \sum_{i=1}^l \lambda_i \right] \end{array} \right)^2 \right\} \\
&\quad \cdot \frac{1}{(2\pi\sigma_\eta^2)^{\frac{n-1}{2}}} \exp \left\{ -\frac{1}{2\sigma_\eta^2} \sum_{t=2}^n \left[ (\alpha_t - \mu_\alpha) - \sum_{i=1}^p \phi_i (\alpha_{t-i} - \mu_\alpha) \right]^2 \right\} \quad (\text{a.3})
\end{aligned}$$

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  $\Omega \sim f_N(\bar{\mu}_\Omega, \bar{V}_\Omega)$ , where

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

From (1.1), (1.3) and (1.2), the regression of  $\Omega$  are with the following equation

$$\begin{aligned}y_t - \exp(\alpha_t)y_{t-1} - \sum_{i=1}^l \lambda_i \Delta y_{t-i} \\ = \gamma [1 - \exp(\alpha_t)] + \delta \left[ t - \exp(\alpha_t)t + \exp(\alpha_t) - \sum_{i=1}^l \lambda_i \right] + \varepsilon_t\end{aligned}\quad (\text{a.4})$$

If we take left hand side of (a.4) as  $Y$

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

and the right hand side of (a.4) as  $X$

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

where  $t = 2, \dots, n$ , Regression Equation (a.4) became a simple linear regression equation

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

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

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

$$\begin{aligned}\bar{V}_\lambda &= (V_\lambda^{-1}h_\varepsilon^{-1} + X^{*'}X^*)^{-1}h_\varepsilon^{-1} \\ &= (V_\lambda^{-1} + X^{*'}X^*h_\varepsilon)^{-1} \\ \bar{\mu}_\lambda &= \bar{V}_\lambda \left( V_\lambda^{-1}h_\varepsilon^{-1}\underline{\mu}_\lambda + X^{*'}X^*\hat{\mu}_\lambda \right) \\ \hat{\mu}_\lambda &= (X^{*'}X^*)^{-1}X^{*'}Y^*\end{aligned}$$

The regression of  $\lambda$  is with the following equation:

$$(y_t - \gamma - \delta t) - \exp(\alpha_t) [y_{t-1} - \gamma - \delta t + \delta] = \sum_{i=1}^l \lambda_i (\Delta y_{t-i} - \delta) + \varepsilon_t \quad (\text{a.5})$$

if we denote

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

$$X^* = (\Delta y_{t-1} - \delta, \dots, \Delta y_{t-i} - \delta)$$

where  $t = 2, \dots, n$ , Equation (a.5) became

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

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

3. The posterior conditionals of error precision  $h_\varepsilon \sim f_{\Gamma_\varepsilon}(\bar{\alpha}_\varepsilon, \bar{\beta}_\varepsilon)$ . According to the prior density of  $h_\varepsilon$

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

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

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

To calculate the posterior of  $\mu_\alpha$ , collect all the parameters together

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

thus variance:

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



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

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

5. The posterior conditionals of  $\phi_i$

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

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

and variance

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

The derivation for  $\phi_1$  are as follows:

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

Then we can have the variance

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

and the mean

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

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

6. The posterior conditionals for  $h_\eta$

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

7. The remaining conditionals that are needed for the  $\alpha_t$  are described as follows:

The conditional densities for  $\alpha_t$  are nonstandard and given by the following expression

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

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

and  $\vartheta(\phi)$  is a function of  $\phi_i$

$\tau(\alpha_t)$  is a function of  $\mu_\alpha, \phi_i, \alpha_{-t}, p$  and  $t$

Derivations are as follows:

$$\begin{aligned}
p(\alpha_t \mid \theta, \alpha_{-t}, y) &\propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \left[\nu_t - \sum_{i=1}^l \lambda_i \Delta \nu_{t-i} - e^{\alpha_t} \nu_{t-1}\right]^2\right\} \\
&\quad \cdot \exp\left\{-\frac{1}{2\sigma_\eta^2} \sum_{t=2}^n \left[(\alpha_t - \mu_\alpha) - \sum_{i=1}^p \phi_i (\alpha_{t-i} - \mu_\alpha)\right]^2\right\}
\end{aligned}$$

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

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

when  $p = 1$

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

For  $t$  from 2 to  $n - 1$

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

where

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

and

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

For  $t = n$

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

where

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

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 it in Tables 15 and 16, which are also available in Jones and Marriott (1999)

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

## B EVALUATION OF THE LIKELIHOOD

According to Equation (1.4),  $y_t$  is a function of latent variable  $\alpha_t$ , thus  $\ln p(y_t | \theta^*, F_{t-1})$  evolves calculation of

$$\begin{aligned} p(y_t | \theta^*, F_{t-1}) &= \int p(y_t, \alpha_t | \theta^*, F_{t-1}) d\alpha_t \\ &= \int p(y_t | \alpha_t, \theta^*, F_{t-1}) p(\alpha_t | \theta^*, F_{t-1}) d\alpha_t \end{aligned} \tag{b.1}$$

As  $\alpha_t$  is non-observable, the exact integrals are hard to obtain. However, with the help of Monte Carlo averaging  $p(y_t | \alpha_t, \theta^*, F_{t-1})$  over the large sample of draws of  $\alpha_t^1, \dots, \alpha_t^M$  from  $p(\alpha_t | \theta^*, F_{t-1})$ , we could have an approximation of  $p(y_t | \theta^*, F_{t-1})$  from the following:

$$p(y_t | \theta^*, F_{t-1}) \simeq \frac{1}{M} \sum_{g=1}^M p(y_t | \alpha_t^{(g)}, \theta^*, F_{t-1}) \quad (\text{b.2})$$

However, a sample of size  $M$  ( $\alpha_t^1, \dots, \alpha_t^M$ ) from  $p(\alpha_t | \theta^*, F_{t-1})$  can be obtained using a sample ( $\alpha_{t-1}^1, \dots, \alpha_{t-1}^M$ ) from  $p(\alpha_{t-1} | \theta^*, F_{t-1})$  as

$$p(\alpha_t | \theta^*, F_{t-1}) = \int p(\alpha_t | \alpha_{t-1}, \theta^*, F_{t-1}) p(\alpha_{t-1} | \theta^*, F_{t-1}) d\alpha_{t-1} \quad (\text{b.3})$$

An Auxiliary Particle Filter (APF) method introduced in Pitt and Shephard (1999) is applied here to get samples from  $p(\alpha_{t-1} | \theta^*, F_{t-1})$ .

The algorithm to evaluate the likelihood can be described as:

**Algorithm: Estimate the log Likelihood for the marginal likelihood using Auxiliary Particle Filter**

First, at time  $t$ , we call the lags of  $\alpha_t$  as  $\underline{\alpha}_t = (\alpha_{t-1}, \dots, \alpha_{t-p})'$ . The lags of  $\alpha_2$  denoted as  $\underline{\alpha}_2^{(g)} : g = 1, \dots, M$ , which are the initial values that can be either set as a  $M \times p$  zeros matrix or a sample of  $M$  draws  $(\alpha_2^{(1)}, \dots, \alpha_2^{(M)})'$  from the conditional prior  $p(\alpha_2 | \theta^*)$ . In our empirical application,  $\underline{\alpha}_2^{(g)} : g = 1, \dots, M$  is set as a  $M \times p$  zeros matrix and  $M$  is set as 3,000.

1.  $t$  starts from 2.

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

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

Note that  $\alpha_t^{\Delta(g)}$  is a draw 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:

$$\hat{p}(y_t | \theta^*, F_{t-1}) = \frac{1}{M} \sum_{g=1}^M p(y_t | \theta^*, \alpha_t^{\Delta(g)}, F_{t-1}) \quad (\text{b.4})$$

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

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

$\lambda_g$  are the first-stage weights. Get  $R$  draws  $(k_1, \dots, k_R)$  from the discrete distribution defined on the integers  $(1, \dots, M)$  with probabilities  $\lambda_1, \dots, \lambda_M$ , where  $R$  is set as large as  $5M$ . This step is to get  $R$  samples from the importance function  $g(\alpha_t, k | \theta^*, F_t)$  by simulating the index with probability  $\lambda_g$ . 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  $g(\cdot)$

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

to get a sample of  $R$  draws.

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

$$\alpha_t^{*(r)} \sim f_N(\underline{\alpha}_t^{k_r}, \phi, \sigma_\eta^{2*}) \quad (\text{b.5})$$

Note that  $(\alpha_t^{*(r)} : r = 1, \dots, R)$  and  $(k_r : r = 1, \dots, R)$  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  $(\alpha_t^{*(r)} : r = 1, \dots, R)'$   $M$  times with probabilities  $\pi_r$  defined as:

$$\begin{aligned} w_r^* &= \frac{p(y_t | \theta^*, F_{t-1}, \alpha_t^{*r})}{p(y_t | \theta^*, F_{t-1}, \hat{\alpha}_t^{k_r})} \\ \pi_r &= \frac{w_r^*}{\sum_{r=1}^M w_r^*} \end{aligned} \quad (\text{b.6})$$

$\pi_r$  is the second-stage weights. Then, the resampled  $M \times 1$  vector, which contains values  $(\alpha_t^{(1)}, \dots, \alpha_t^{(M)})'$  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  $g(\cdot)$ . Stacking this sampled  $(\alpha_t^{(1)}, \dots, \alpha_t^{(M)})'$  on  $\underline{\alpha}_t^{(g)} : g = 1, \dots, M$ . We have the updated lags of  $\underline{\alpha}_{t+1}^g = (\alpha_t^{(g)}, \alpha_{t-1}^{(g)}, \dots, \alpha_{t-(p-1)}^{(g)}) : g = 1, \dots, M$

5. Fix  $t = t + 1$  go to step 1(a) to get  $\alpha_{t+1}^{\Delta(g)}$  from  $\underline{\alpha}_{t+1}^g$  in step 4. Note that  $\alpha_{t+1}^{\Delta(g)}$  are samples from  $p(\alpha_{t+1}^{\Delta(g)} | \theta^*, F_t, \underline{\alpha}_{t+1}^g)$  and following 1(b) to get  $\hat{p}(y_{t+1} | \theta^*, F_t)$ <sup>5</sup>, until  $t = n$ .
6. Finally, the estimate of the log likelihood is

$$\log \hat{p}(y | \theta^*) = \sum_{t=2}^N \log \hat{p}(y_t | \theta^*, F_{t-1}) \quad (\text{b.7})$$

---

<sup>5</sup>An estimate of  $p(y_{t+1} | \theta^*, F_t)$  is given by:

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

## 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*, Springer-Verlag.
- [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] Geweke, J (1989): “Bayesian Inference in Econometric Models using Monte Carlo Integration”, *Econometrica*, 57, 1317-1339.
- [8] Gil-Alana, L.A. and P.M.Robinson. (1997): “Testing of Unit Root and Other Non-stationary Hypotheses in Macroeconomic Time Series”, *Journal of Econometrics*, 80, 241-268.
- [9] Granger, C. W. J. and N. R. Swanson. (1997): “An Introduction to Stochastic Unit-Root Processes”, *Journal of Econometrics*, 80, 35-62.
- [10] 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.
- [11] Jones, C. R. and M. J. Marriott. (1999): “A Bayesian Analysis of Stochastic Unit Root Models”, *Bayesian Statistics*, 6, 785-794.
- [12] Judge, G. G. (1985): *The Theory and Practice of Econometrics*, Wiley.
- [13] Kapetanios, G; Y. Shin and A. Snell. (2003): “Testing for a Unit Root in the Nonlinear STAR Framework”, *Journal of Econometrics*, 112, 359-379.
- [14] Kass, R. E. and A. E. Raftery. (1995): “Bayes Factors”, *Journal of the American Statistical Association*, 90, 773-795.
- [15] Kim, J.Y. (2000): “Detection of Change in Persistence of a Linear Time Series”, *Journal of Econometrics*, 95, 97-116.
- [16] –, and R. B. Amador. (2002/8): “Corrigendum to Detection of Change in Persistence of a Linear Time Series”, *Journal of Econometrics*, 109, 389-392.
- [17] 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.

- [18] –, 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.
- [19] –, and S. M. Potter. (1999/2): “Bayes Factors and Nonlinearity: Evidence from Economic Time Series”, *Journal of Econometrics*, 88, 251-281.
- [20] –, (2003): *Bayesian Econometrics*, Wiley.
- [21] 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.
- [22] Nelson, C. R. and C. I. Plosser. (1982): “Trends and Random Walks in Macroeconomic Time Series”, *Journal of Monetary Economics*, 10, 139-162.
- [23] 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.
- [24] Pitt, M. K. and N. Shephard. (1999): “Filtering Via Simulation: Auxiliary Particle Filters”, *Journal of the American Statistical Association*, 94, 590-599.
- [25] van Dijk, D.; Tervirta T. and P. H. Franses. (2002): “Smooth Transition Autoregressive Models — A Survey of Recent Developments”, *Econometric Reviews*, 21, 1-47.
- [26] Watson, M. and J.H. Stock. (2007/1): “Why has inflation become harder to forecast?”, *Journal of Money Credit and Banking*, 39, 3-33.

Table 4: log Marginal Likelihood with a GSTUR class of Models

	$\gamma \neq 0, \delta \neq 0$	$\gamma \neq 0, \delta = 0$	$\gamma = 0, \delta \neq 0$	$\gamma = \delta = 0$
(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 5: Estimates: GSTUR with U.K./U.S.Real Exchange Rates

	<i>Prior</i>		<i>Posterior</i>						
	<i>Mean</i>	<i>St.Dev</i>	<i>Mean</i>	<i>St.Dev</i>	<i>CD</i>	<i>NSE<sub>.15</sub></i>	<i>Median</i>	<i>95%Posterior Band</i>	
$\mu_\alpha$	ln 0.9	0.1 <sup>2</sup>	-0.0211	0.0510	0.0204	0.0006	-0.0090	-0.1330	0.0276
$\sigma_\eta^2$	-	-	0.0002	0.0002	-0.1093	0.0000	0.0001	0.0000	0.0006
$\sigma_\varepsilon^2$	-	-	0.0507	0.0020	-0.2337	0.0000	0.0506	0.0475	0.0540
$\phi_1$	†	†	0.4375	0.4855	-0.3223	0.0112	0.5488	-0.5100	0.9812
$\lambda_1$	0	10 <sup>6</sup>	0.9461	0.0555	1.0901	0.0018	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 Table (1) for description

Table 6: Full Sample: Forecast of Inflation Rates with GSTUR p=1 lag=7, with Cons but no trend

	TrueV	Density forecast				95% percentile		
		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 7: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with Trend but no constant

	TrueV	Density forecast				95% percentile		ML
		mean	mode	median	var	2.5%	97.5%	
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 8: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with Cons and Trend

	TrueV	Density forecast				95% percentile		ML
		mean	mode	median	var	2.5%	97.5%	
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 9: Full Sample:Forecast of Inflation Rates GSTUR p=1 lag=7, with no constant and no trend

	TrueV	Density forecast				95% percentile		ML
		mean	mode	median	var	2.5%	97.5%	
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 10: Full Sample:Forecast of Inflation Rates with RW

	TrueV	Density forecast			var	95% percentile		MLint
		mean	mode	median		2.5%	97.5%	
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 11: Full Sample: MSFE of Statistical Forecasting Models

	$l = 0$	$l = 1$	$l = 4$	$l = 5$	$l = 7$
GSTUR: Constant no trend	0.1331	0.0706	0.058	0.0618	0.0578
GSTUR: Trend no constant	0.1302	0.0746	0.0553	0.0609	0.0605
GSTUR: Constant and Trend	0.1386	0.0792	0.0636	0.0627	0.0629
GSTUR: No constant no Trend	0.1539	0.0863	0.0683	0.073	0.0692
RW: 0.1535					

Table 12: Forecast of Inflation Rates in Small Sample GSTUR p=1 lag=4 with NO CT

	TrueV	Density forecast			var	95% percentile		ML
		mean	mode	median		2.5%	97.5%	
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 13: Forecast of Inflation Rates in Small Sample with a RW Model

	TrueV	Density forecast			var	95% percentile		ML
		mean	mode	median		2.5%	97.5%	
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

Table 14: Small Sample: MSFE of Statistical Forecasting Models

RW	GSnoCT: $l = 5$	GSnoCT: $l = 1$	GSnoCT: $l = 4$	AR(5)
0.1335	0.1268	0.1021	0.0987	0.1879

Table 15: Appendix: Functions for Sampling  $\alpha$  (a)

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

Table 16: Appendix: Functions for Sampling apha(b)

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

---

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

---

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

## SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Volatility Investing with Variance Swaps" by Wolfgang Karl Härdle and Elena Silyakova, January 2010.
- 002 "Partial Linear Quantile Regression and Bootstrap Confidence Bands" by Wolfgang Karl Härdle, Ya'acov Ritov and Song Song, January 2010.
- 003 "Uniform confidence bands for pricing kernels" by Wolfgang Karl Härdle, Yarema Okhrin and Weining Wang, January 2010.
- 004 "Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model" by Nikolaus Hautsch and Fuyu Yang, January 2010.
- 005 "The Impact of Macroeconomic News on Quote Adjustments, Noise, and Informational Volatility" by Nikolaus Hautsch, Dieter Hess and David Veredas, January 2010.
- 006 "Bayesian Estimation and Model Selection in the Generalised Stochastic Unit Root Model" by Fuyu Yang and Roberto Leon-Gonzalez, January 2010.