Bayesian Inference and Forecasting in the Stationary Bilinear Model

ROBERTO LEON-GONZALEZ † and FUYU YANG ‡

†*National Graduate Institute for Policy Studies,7-22-1 Roppongi, Minato-ku, Tokyo, Japan, 106-8677*
(e-mail: email: rlg@grips.ac.jp)
‡*School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK*
(e-mail: fuyu.yang@uea.ac.uk)

Abstract

A stationary bilinear (SB) model can be used to describe processes with a time-varying degree of persistence that depends on past shocks. The SB model can be used to model highly persistent macroeconomic time series such as inflation. This study develops methods for Bayesian inference, model comparison, and forecasting in the SB model. Using U.K. inflation data, we find that the SB model outperforms the random walk and first order autoregressive AR(1) models, in terms of root mean squared forecast errors for the one-step-ahead out-of-sample forecast. In addition, the SB model is superior to these two models in terms of predictive likelihood for almost all of the forecast observations.

JEL classification numbers: C5, C11, C32

Keywords: Stationary bilinear model, Markov chain Monte Carlo, model comparison

1 Introduction

The class of bilinear processes was first proposed by Granger and Andersen (1978a) and was found to be able to, as stated in Raeburn et al. (1995), 'approximate any nonlinear model to an arbitrary degree of accuracy over a finite time interval'. Brunner and Hess (1995) point out that the bilinear model's capacity to approximate any well-behaved nonlinear relationship is analogous to the ability of an autoregressive moving average (ARMA) model to approximate well-behaved linear relationships. Bilinear models have been successfully applied to analyze macroeconomic and financial series to capture data non-linearity; see, for example, Charemza et al. (2005), Byers and Peel (1995), and Hristova (2005). In this study, we focus on making Bayesian inference in a stationary bilinear (SB) model with a single bilinear term, and on comparing the forecasting capacity of the nonlinear SB model, the linear random walk (RW) model, and a lower order autoregressive (AR) model.

The SB model is specified as the following:

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

where b is the bilinear term and $\varepsilon_t \sim i.i.d.N\left(0, \sigma_\varepsilon^2\right)$. Granger and Andersen (1978a) point out a second order stationary condition for the SB model is

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

where $a,\ b \neq 0$. Along with equation (2), Sesay and Subba Rao (1988) and Kim et al. (1990) established the following necessary restrictions (equations 3 –5) to meet the assumption that the first four moments of $\{y_t\}_{t=1}^T$ are finite.

$$|a| < 1, \tag{3}$$

$$\left|a^3 + 3ab^2\sigma_\varepsilon^2\right| < 1, \tag{4}$$

and

$$a^4 + 6a^2b^2\sigma_\varepsilon^2 + 3b^4\sigma_\varepsilon^4 < 1. \tag{5}$$

In this study, we focus on making Bayesian inference in a SB model, where a, b, and σ_ε^2 jointly meet the moment restrictions in equations (2)–(5).

The dynamics in a SB process are driven by ε_t, and the persistence parameter, $a + b\varepsilon_{t-1}$, is driven by the past shock ε_{t-1}. Whereas, in O'Reilly and Whelan (2005), the persistence parameter is defined as 'the sum of the coefficients on the lagged dependent variable'. A non-zero bilinear term b together with the error term ε_{t-1}, therefore, would induce a time-varying persistence that changes corresponding to the lagged shocks ε_{t-1}.

According to equation (1), the one-step-ahead forecast of y_{t+1} using the SB model is as follows:

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

which indicates that a big shock in the system may affect both the expectations of future values and induce changes in the series' persistence. Moreover, if a series is modelled with a SB process, all the past shocks $\{\varepsilon_i\}_{i=1}^{t}$ are inevitably propagated via equation (1) to affect the one-step-ahead forecast $E\left(y_{t+1}\right)$. If $E\left(y_{t+1}\right)$ is constructed in such a way that the bilinear term b and ε_t are neglected, the forecast is distorted; see Charemza et al. (2005) for empirical illustrations. For instance, if b is misspecified as 0, the SB model is equivalent to a linear AR(1) model:

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

Therefore, the forecast is distorted by the amount, $b\varepsilon_t y_t$, if the forecasting model is misspecified as an AR(1) model.

The SB process is very useful for empirical applications. However, as pointed out in Brunner and Hess (1995), the estimation of a SB model could be problematic, when at least one of the four moment conditions in equations (2)–(5) are close to being violated.

Brunner and Hess (1995) simulate 10,000 series of data using a first order bilinear process (BL(1,0,1,1)) with a few sets of 'true' values, where the first moment condition (which is also one of the stationary conditions) is close to being violated. They demonstrate that the expected negative log-likelihood function will be characterized with a long narrow spike under the 'true' values. Therefore, because of this characteristic in the expected log-likelihood function, the estimates of the bilinear parameters from a standard optimization routine, e.g. maximum likelihood method, tend to bias away from the 'true' values. However, if the

moment conditions are easily satisfied, the well behaved expected log-likelihood function has a global optimum located over the 'true' parameter values, and the global optimum can be found easily.

In this study, we make Bayesian inference in a SB model, which has a single bilinear term. The Bayesian estimation method for the SB model, and the model comparison method for non-nested models that we propose may overcome some of the difficulties highlighted in the literature. The SB model specification satisfies the stationary conditions, which makes it distinct from many other bilinear model specifications that have been investigated in the literature, such as in Feng et al. (2013), Bibi and Lessak (2009), Charemza et al. (2005), and Chen (1992a). We aim to develop an efficient sampling algorithm, which enables us to simulate posterior distributions for all parameters of interest jointly satisfying the stationary restrictions.

Chen (1992a) has also proposed to estimate the bilinear models using Bayesian techniques. In Chen (1992a), the priors were elicited without giving the stationary conditions a special consideration. However, in the Bayesian framework, a posterior simulator is normally constructed by combining the prior and the likelihood function. Therefore, the elicited priors should reflect a 'priori' distributional belief in the parameters of interest to a certain extent, if not fully. In particular, considering the characteristics of the expected log-likelihood function of the bilinear models in Brunner and Hess (1995), the prior elicition is crucial for making Bayesian inference in the SB model. The Gibbs sampling algorithm proposed by Chen (1992a) is efficient using the untruncated elicited priors because the data simulated for applications in Chen (1992a) can easily satisfy the stationary conditions. However, if a bilinear data generating process (DGP) does not easily satisfy the first moment condition, the Gibbs sampling algorithm proposed in Chen (1992a), and forecasting method in Chen (1992b), will not be applicable.

Therefore, the sampling algorithm proposed by Chen (1992a) is not suitable for bilinear model specifications with the first moment condition close to being violated. However, for macroeconomic modellers, such bilinear model specifications are more appealing and intuitive, because this type of underlying process allows the data to be highly persistent, and it also allows the persistence to be correlated with shocks in previous periods.

In this study, with an application to U.K. inflation, we compare the in-sample

fit of three non-nested models, the SB model, the RW model, and a lower order AR model using the likelihood and the marginal likelihood. To compare the out-of-sample forecasting capacity of competing models, besides evaluating the root mean squared forecast errors (RMSFE), the comparison is also carried out using the predictive likelihood; see Geweke and Amisano (2010).

The remainder of the paper is organized as follows. Section 2 presents methods in Bayesian inference in the SB model specified in equation (1). Section 3 uses U.K. inflation data to estimate the SB model and to compare it to other models in terms of forecasting accuracy at different horizons. Section 4 concludes.

2 Bayesian Inferences in the SB Model

We start by introducing some notation: a time series with a sample size of N is denoted as $\mathbf{y} = (y_1, \cdots, y_n)'$, and we assume the first observation y_1 is the initial observation. The error disturbances ε_t for $t = 1, \cdots, n$ are $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$. Following the recommendations in Charemza et al. (2005), the initial disturbance value is set as $\varepsilon_1 = 0$. We denote the error precision $h_\varepsilon = \sigma_\varepsilon^{-2}$ and collect all parameters of interest in θ, where $\theta = (a, b, h_\varepsilon)'$. Let $N(\mu, V)$ denote the normal distribution with mean μ and variance V. Therefore, the normal density is denoted by $f_N(\mu, V)$. Suppose a random variable x follows a gamma distribution, $x \sim \Gamma(\alpha, \beta)$, where α is the shape parameter and β is the scale parameter, then the gamma density is $f_\Gamma(\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$.

A truncated normal prior for a is elicited, $p(a) = f_N\left(\overline{\mu}_a, \overline{V}_a\right) I_{[|a|<1]}$, where $I_{[A]}$ is the indicator function indicating that the moment restriction in equation (3) is satisfied. By choosing $\overline{\mu}_a = 0$ and $\overline{V}_a = 10^4$, the truncated normal prior is roughly equivalent to a uniform prior in the interval that verifies the restriction. We elicit a gamma prior for the error precision h_ε as $p(h_\varepsilon \mid a) = f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$. Note that the typical improper non-informative prior ($p(h_\varepsilon \mid a) \propto |h_\varepsilon|^{-1}$) corresponds to $p(h_\varepsilon \mid a) = f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ with $\underline{\alpha}_\varepsilon = 0$ and $\underline{\beta}_\varepsilon = \infty$. To make the prior for h_ε as non-informative as possible, yet ensure it is a proper prior, we choose $\underline{\alpha}_\varepsilon = 1$. By choosing $\underline{\alpha}_\varepsilon = 1$, the distribution of $\sigma_\varepsilon^2 = h_\varepsilon^{-1}$ is an inverted gamma with one degree of freedom. Therefore, it has no prior moments with infinite mean and variance, which implies great uncertainty. We may try a range of values for $\underline{\beta}_\varepsilon$ for prior robustness analysis. The smaller $\underline{\beta}_\varepsilon$ is, the greater prior weight is given to

large values of σ_ε^2, and this results in a greater chance that y_t is large in absolute value. We have elicited $\underline{\beta}_\varepsilon = 0.01$ as a reasonable value and try departures from this value as well in a sensitivity analysis.

Using equation (1) recursively, we can write

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

for $t = 2, \cdots, n$, so that y_t can be written as

$$y_t = a f_1(t, b) + f_2(t, b) + \varepsilon_t, \tag{6}$$

where

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

and

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

Note that $f_2(t, b) = 0$ when $t = 2$. The corresponding derivations can be found in Appendix A. Denote $F_{t-1} = (y_1, \cdots, y_{t-1})'$. According to equation (6), the likelihood function of $p(y_t \mid \theta, F_{t-1})$ for $t \geq 2$ is as follows:

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

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

$$p(\mathbf{y} \mid \theta) = \frac{h_\varepsilon^{\frac{N-1}{2}}}{(2\pi)^{\frac{N-1}{2}}} \exp\left\{ -\frac{h_\varepsilon}{2} \sum_{t=2}^{N} \left[y_t - a f_1(t, b) - f_2(t, b) \right]^2 \right\}. \tag{7}$$

The density function of the prior, which incorporates the moment restrictions, is

$$p(a, b, h_\varepsilon) = p(b \mid h_\varepsilon, a)\, p(h_\varepsilon \mid a)\, p(a)\, I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1\right]} I_{[M]}, \tag{8}$$

where M is the region in which the inequalities (4) and (5) are satisfied. From equation (7), we are not able to elicit a prior for b such that the conditional posterior follows a standard distribution. Therefore, for simplicity, we assume b to be uniformly distributed within a region such that the moment condition in

equation (2) is satisfied. With the derived likelihood in equation (7) and the joint prior in equation (8), the joint posterior is given by the Bayes theorem as

$$\begin{aligned} p\left(a, b, h_\varepsilon \mid y\right) &\propto p\left(y \mid a, b, h_\varepsilon\right) p\left(a, b, h_\varepsilon\right) \\ &\propto p\left(y \mid a, b, h_\varepsilon\right) p\left(b \mid h_\varepsilon, a\right) p\left(h_\varepsilon \mid a\right) p(a) I_{\left[a^2+b^2 h_\varepsilon^{-1}<1\right]} I_{[M]}. \end{aligned}$$

The conditional posterior of h_ε is a truncated gamma and that of a is a truncated normal.

The biggest challenge in developing the sampling algorithm is to take random draws from the truncated posteriors. Because the truncation area could be far in the tail of the distribution, to draw from these truncated distributions, we use the mixed rejection algorithm proposed by Geweke (1991). The following Gibbs sampling scheme uses the conditional posterior densities outlined in Appendix A to get draws from the posterior.

Gibbs sampling for a, b, **and** h_ε

1. Give initial values to a and b.
2. Repeat (a)–(c) S times and discard the first S_0 draws.
   (a) Sample h_ε from $h_\varepsilon \mid a, b \sim \Gamma\varepsilon\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right) \cdot I_{[h_\varepsilon > b^2/(1-a^2)]} I_{[M]}$ using mixed rejection sampling.
   (b) Sample a from $a \mid b, h_\varepsilon \sim N\left(\overline{\mu}_a, \overline{V}_a\right) \cdot I_{\left[|a|<\sqrt{1-b^2 h_\varepsilon^{-1}}\right]} I_{[M]}$ using mixed rejection sampling.
   (c) Sample b from a region, where $b \mid a, h_\varepsilon \sim I_{\left[|b|<\sqrt{(1-a^2)h_\varepsilon}\right]} I_{[M]}$ using the Griddy–Gibbs sampling algorithm; for example, see Ritter and Tanner (1992).

The averages of the draws from the Gibbs sampling are estimates of the posterior means. Convergence diagnostics can be conducted to examine the efficiency of the proposed Markov chain Monte Carlo (MCMC) algorithm.

3 Application to U.K. Inflation Rates

Because the SB specification accommodates both nonlinearity and time varying features, the SB model can be a natural candidate for the inflation modelling.

Forecasting inflation has been a main interest of macroeconomic modellers. In particular, econometricians focus on the behaviour of inflation persistence, which is closely related to the monetary regime. It is suggested in Osborn and Sensier (2009) and O'Reilly and Whelan (2005), that the persistence of inflation may have changed in the last three decades in the U.K. and Euro area due to substantial changes in the monetary regime over time. We aim to develop an efficient sampling algorithm for the SB model and apply it to U.K. inflation.

To illustrate empirically the efficiency of the proposed MCMC algorithm, we apply the quarterly U.K. inflation rate to the SB model. The data set has 201 observations covering the period from quarter 1-1957 to quarter 1-2007. The data is collected for headline U.K. inflation rates, which are the annual percentage changes in the seasonally unadjusted Retail Price Index (RPI). The RPI is all inclusive and includes mortgage payments. In fact, under the inflation targeting policy in the U.K., the target range for inflation is actually measured using the RPI after excluding mortgage payments (i.e., the RPIX). Two important events occurred in the sample periods: In 1992, a new framework for monetary policy commenced following the U.K.'s exit from the European Community's Exchange Rate Mechanism. In 1997, the central bank was granted operational independence over monetary policy, and inflation was targeted.

Model Estimation

We fit the data with the first order SB model in equation (1) using the MCMC algorithm proposed in Section 2. Table (1) provides the estimation results using the Gibbs sampler from $12,000$ iterations. The first $2,000$ draws are discarded. For prior sensitivity analysis, we choose two different priors of a: a tight prior with $\underline{V}_a = 0.01$, and a flat prior with $\underline{V}_a = 10^4$. A Griddy–Gibbs sampler algorithm is applied to sample b. Because the sampling algorithm is efficient and the computational cost of having numerous grids is small, we set the grid number as $1,500$. From table (1), all absolute convergence diagnostic (CD) values are smaller than 1.96. Therefore, we do not find evidence that the chain does not converge. Because there is not much difference in the estimated results when a flat prior or a tight prior is used, the algorithm is not sensitive to elicited priors.

Figure 1 plots the posterior draws and histogram plots. Results from the

Table 1: *Estimation results from the SB model applied to U.K. inflation data. The elicited prior a is* $a \sim N\left(\overline{\mu}_a, \overline{V}_a\right) I_{[|a|<1]}$, *where* $\overline{\mu}_a = 0.98$. *For prior sensitivity analysis,* $\overline{V}_a = 10^4$ *and* $\overline{V}_a = 0.01$ *are elicited for comparison. The prior for* h_ε *is* $h_\varepsilon \sim \Gamma\varepsilon\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, *where* $\underline{\alpha}_\varepsilon = 1$ *and* $\underline{\beta}_\varepsilon = 0.01$. *The table reports posterior means, modes, and medians together with standard deviations. CD represents convergence diagnostic values.*

	$a \sim N\left(\underline{\mu}_a, \underline{V}_a\right) I_{[\|a\|<1]}$ *with* $\underline{\mu}_a = 0.98$, $\underline{V}_a = 10^4$				
$\underline{\alpha}_\varepsilon = 1, \underline{\beta}_\varepsilon = 0.01$	*Mean*	*Mode*	*Median*	*St.Dev*	*CD*
a	0.9752	0.9793	0.9767	0.0127	−0.1437
b	0.0468	0.0481	0.0474	0.0032	0.7677
σ_ε^2	1.2964	1.2482	1.2882	0.1332	−1.0643
	$a \sim N\left(\underline{\mu}_a, \underline{V}_a\right) I_{[\|a\|<1]}$ *with* $\underline{\mu}_a = 0.98$, $\underline{V}_a = 0.01$				
$\underline{\alpha}_\varepsilon = 1, \underline{\beta}_\varepsilon = 0.01$	*Mean*	*Mode*	*Median*	*St.Dev*	*CD*
a	0.9756	0.9783	0.9772	0.0127	0.4667
b	0.0467	0.0484	0.0474	0.0032	−0.5932
σ_ε^2	1.2957	1.2755	1.2870	0.1293	−0.6484

Bayesian MCMC approach indicate that a is close to unity and that the bilinearity term b is non-negligible. This non-zero b indicates that past shocks have an impact on current inflation. Figure 2 plots the changing persistence, filtered shocks, and the data. Not surprisingly, there are big shocks in the 1970s and early 1980s. As a result, there is a big variation in the inflation persistence between 1972 and 1982. Correspondingly, the changes in inflation persistence after 1991 are smaller relative to those before early 1980s. Our results confirm the finding in the literature that inflation persistence has declined after 1980s, and it has been time-varying.

Model Comparison

Applying the Phillips - Perron (PP) and augmented Dickey - Fuller (ADF) tests to the U.K. inflation series, we cannot reject the null hypothesis that a series has a unit root at all significance levels. In a Bayesian framework, the model uncertainties can be quantified. In other words, we wish to establish whether the underlying process follows a RW process, a lower order process, i.e. an AR(1) process, or an SB process.

We apply three methods to compare the RW, AR(1), and SB models. The first method is to calculate the log-likelihood using the fitted parameter estimates.

Figure 1: Plots of the MCMC draws. The top panel plots the posterior draws (from left to right) for a, b, and σ_ε^2. The bottom panel plots the simulated posterior distributions (from left to right) of a, b, and σ_ε^2.

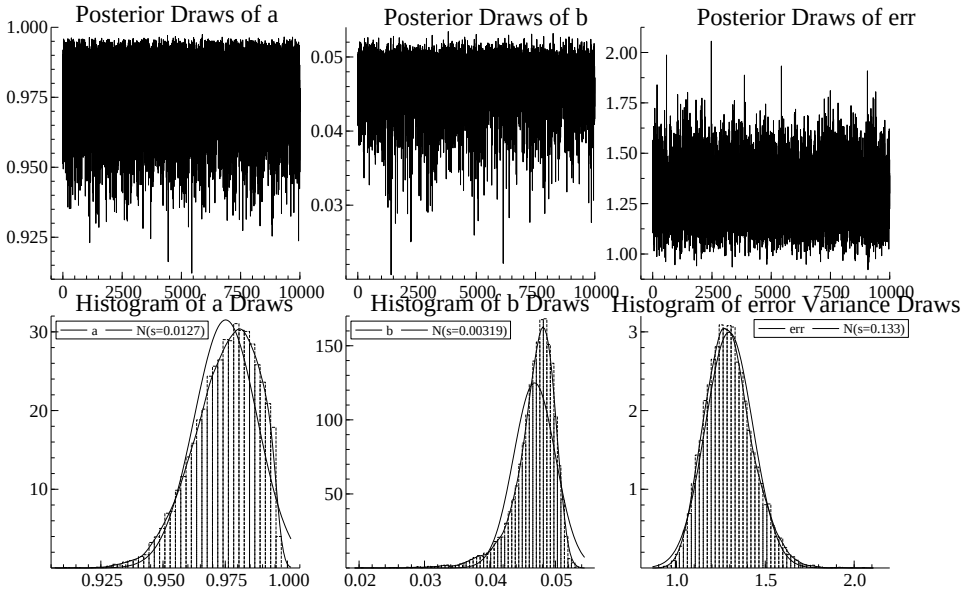
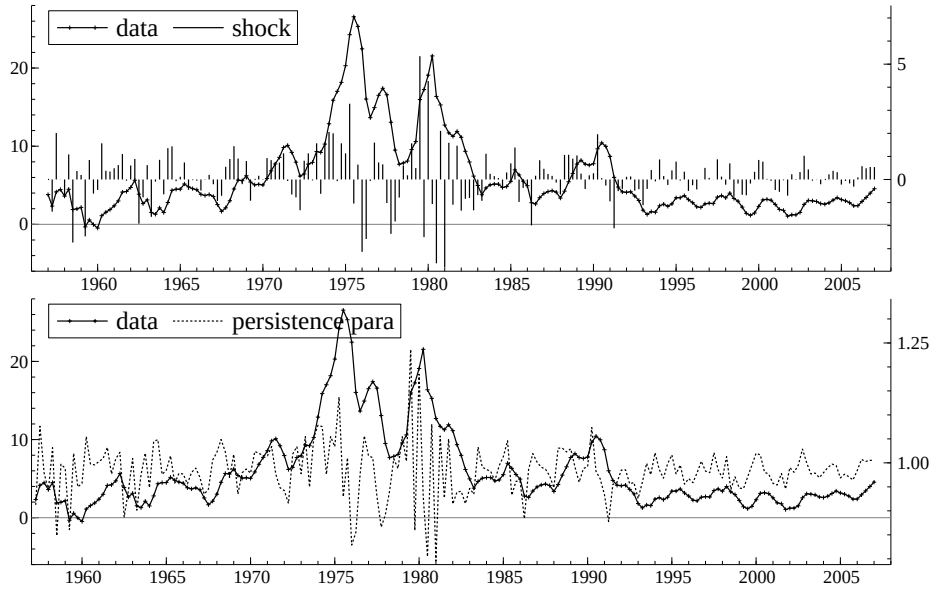


Figure 2: From top to bottom: Inflation data with corresponding filtered shocks $\widehat{\varepsilon}_t$; Inflation data and time-varying persistence parameter $\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_{t-1}\right)$. The entire sample period is from quarter 2-1957 to quarter 1-2007; $t = 2, \cdots, 201$



The second method is to calculate the marginal likelihood using a Gelfand–Dey method; see Gelfand and Dey (1994). The marginal likelihood of the RW model

can be calculated analytically by integrating out the only unknown parameter σ_ε^2. The third method is to calculate the predictive likelihood proposed by Geweke (2001), which is less sensitive to the choice of priors relative to the second method using the marginal likelihood.

In Geweke (2001), the predictive likelihood is calculated by integrating the likelihood function over the posterior distribution of the unobservables at the time the prediction is made. In line with the idea that a model is as good as its predictions, the predictive likelihood provides an intuitive tool to compare across models with the Bayesian approach.

Denote as θ_{SB} the parameters in M_{SB}, the SB model. Denote as $\mathbf{y}_{t-1}^o$ all observations available up to $t-1$. The one-step-ahead predictive likelihood, evaluated at time t, is defined as

$$PL_{SB}(t) = p\left(y_t^o \mid \mathbf{y}_{t-1}^o, M_{SB}\right) = \int_{\ominus_A} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, \theta_{SB}\right) p\left(\theta_{SB} \mid \mathbf{y}_{t-1}^o\right) d\theta_{SB},$$

which can be approximated by

$$S^{-1} \sum_{s=1}^{S} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, \theta_{SB}^{(s)}\right),$$

where $\theta_{SB}^{(s)}$ are the MCMC draws from the posterior of θ_{SB} given data up to $t-1$. The predictive likelihood for all observations from $t=2$ up to T is equal to the marginal likelihood and is defined as

$$p\left(\mathbf{y}_t^o \mid M_{SB}\right) = \prod_{t=2}^{T} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, M_{SB}\right),$$

which implies the additive decomposition:

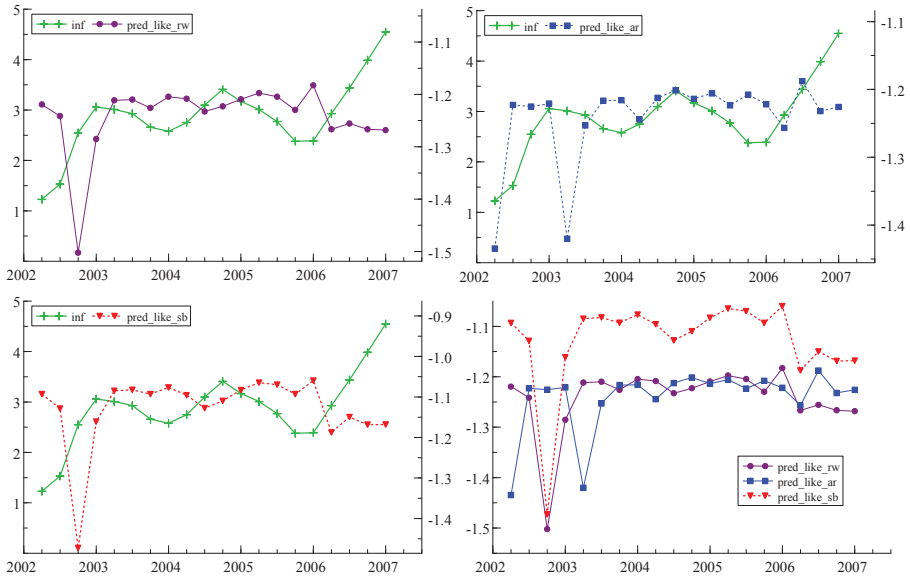
$$\log p\left(\mathbf{y}_t^o \mid M_{SB}\right) = \sum_{t=2}^{T} \log PL_{SB}(t).$$

Because one of the competing models is the RW model, M_{RW}, the log Bayes factor can then be decomposed as

$$\log\left[\frac{p\left(\mathbf{y}_t^o \mid M_{SB}\right)}{p\left(\mathbf{y}_t^o \mid M_{RW}\right)}\right] = \sum_{t=2}^{T} \log\left[\frac{PL_{SB}(t)}{PL_{RW}(t)}\right].$$

As stated in Geweke and Amisano (2010), the decomposition shows how individual observations contribute to the evidence in favor of the nonlinear SB model versus the linear RW model and the AR(1) model.

Figure 3: Reserved 20 observations from quarter 2-2002 to quarter 1-2007. In the top panel from left to right: data and the predictive likelihood of the RW model; data and the predictive likelihood of the AR(1) model. In the bottom panel from left to right: data and the predictive likelihood of the SB model; predictive likelihood of the RW, AR(1), and SB models at each time period.



If we calculate $\log PL_{SB}(t)$ from $t = k+1$ rather than from $t = 2$, the initial k observations are then used as a training sample. The predictive likelihood for the last $T - k$ observations is

$$\log p\left(\mathbf{y}_t^o \mid \mathbf{y}_k^o, M_{SB}\right) = \sum_{t=k+1}^{T} \log PL_{SB}(t).$$

In this study, the initial 181 observations are chosen as the training sample. Then, we can use the log predictive likelihood for the last 20 observations (quarter 2-2002 – quarter 1-2007) as a tool for model comparison. Furthermore, we plot the contribution of each data point in the log predictive likelihood. In this way, we can identify the observations that contribute toward the evidence in favor of the SB model versus the RW and AR(1) models. In table (2), the log likelihoods are calculated by evaluating the log likelihood function at the posterior means of the parameters. We can see that the SB model has the highest likelihood value. Further, the log marginal likelihood of the SB model is the highest among the competing models. However, this result might be sensitive to the selection of the prior.

Table 2: *Model Comparison Results: Calculated log likelihood using the whole sample, log marginal likelihood using the whole sample, and log predictive likelihood of the last* 20 *observations. The competing models are SB, RW, and AR(1).*

	log likelihood	*log marginal likelihood* $p = 0.01$	*log predictive likelihood 181 training sample*
SB flat prior	−308.7578	−253.5415	−22.5708
SB tight prior	−308.7233	−247.5232	−22.5686
RW	−335.5857	−338.0864	−24.8443
AR(1)	−336.5346	−341.1809	−24.8427

The predictive likelihood of the last 20 observations also indicates that the SB model is the best. At the bottom right of Figure 3, how each observation contributes to the predictive likelihood for each of the three competing models is plotted. Comparing the predictive likelihood over multiple horizons for observed data using the three competing models, we can see that the SB model obtains the highest predictive likelihood in most cases, specifically, in 19 out of 20 observations.

Forecasting Exercise

Regarding inflation forecasting, a considerable section of the literature focuses on investigating the forecasting performance of linear and nonlinear econometric models in both univariate and multivariate time series framework, e.g. Barnett et al. (2014), Maheu and Song (2014), Stock and Watson (1999), Stock and Watson (2007), etc. In Canova (2007), bivariate and trivariate models suggested by the economic theory add on marginal predictive accuracy compared with the univariate model, after a horse race between multivariate models and univariate models. Based on the forecasting comparison results, it turns out that the random walk (RW) model or a lower order AR model often provides the smallest RMSFE, and therefore, are hard to beat, see Atkeson and Ohanian (2001).

In this section, with an application to U.K. inflation, we simulate forecast densities by taking advantages of the Bayesian framework. We then compare the forecasting capacity of the SB model with other leading forecasting models, such as lower order AR models, and the RW model.

Granger and Andersen (1978a, p. 74) discussed the invertibility conditions of the bilinear models. Also, in Subba Rao (1981), invertibility conditions for

a generalized bilinear time series models are provided. Using the definition of invertibility in Granger and Andersen (1978b), the SB process is invertible if $|b\sigma_\varepsilon| < 0.606$. Limiting ourselves to a simple case, based on the estimation results from table (1), this nonlinear SB model is not subject to a non-invertibility problem. Therefore, the SB forecasting model can be used to associate the present events with past happenings in a unique manner.

Denoting $A_{t-1} = a + b\varepsilon_{t-1}$, the SB data generating process can be rewritten as

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

Therefore, we can write $y_{t+1} = A_t y_t + \varepsilon_{t+1}$, where $A_t = a + b\varepsilon_t$. The one-step-ahead point forecast of y_{t+1} conditional on the current observation y_t, denoted $\widetilde{y}^{SB}_{t+1|t}$ is

$$\begin{aligned} \widetilde{y}^{SB}_{t+1|t} &= E_t\left[y_{t+1} \mid y_t\right] = E_t\left[A_t y_t + \varepsilon_{t+1} \mid y_t\right] \\ &= \widehat{A}_t y_t + E\left(\varepsilon_{t+1} \mid y_t\right) = \widehat{A}_t y_t, \end{aligned}$$

where $\widehat{A}_t = \widehat{a} + \widehat{b}\widehat{\varepsilon}_t$. The estimates of all past shocks $\widehat{\varepsilon}_i$ with $i = 2, ..., t$ can be retrieved using $\widehat{\varepsilon}_t = y_t - \left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_{t-1}\right) y_{t-1}$, where the initial shock ε_1 is assumed to be 0.

If we denote the two-step-ahead point forecast of y_{t+2} conditional on the current observation y_t as $\widetilde{y}^{SB}_{t+2|t}$, given $\varepsilon_{t+1} \sim N\left(0, \sigma^2_\varepsilon\right)$, then the two-step-ahead forecast $\widetilde{y}^{SB}_{t+2|t}$ is

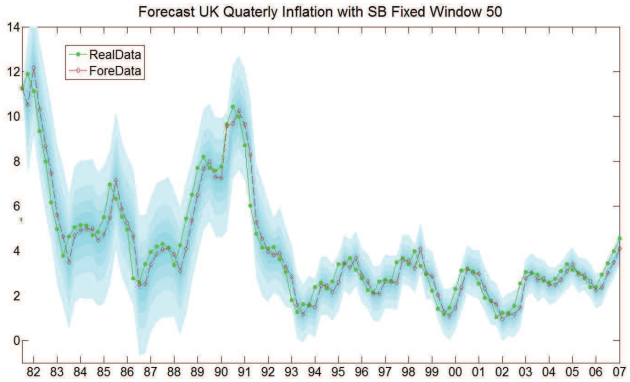
$$\begin{aligned} \widetilde{y}^{SB}_{t+2|t} &= E_t\left[y_{t+2} \mid y_t\right] = E_t E_{t+1}\left[y_{t+2} \mid y_t\right] = E_t E_{t+1}\left[A_{t+1}y_{t+1} + \varepsilon_{t+2} \mid y_t\right] \\ &= E_t E_{t+1}\left[A_{t+1}\left(A_t y_t + \varepsilon_{t+1}\right) \mid y_t\right] = \widehat{a}\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_t\right) y_t + \widehat{b}\widehat{\sigma}^2_\varepsilon. \end{aligned}$$

According to the law of iterated expectations, an h-step-ahead point forecast can be summarized as

$$\begin{aligned} \widetilde{y}^{SB}_{t+h|t} &= E_t\left[y_{t+h} \mid y_t\right] = E_t E_{t+1} \ldots E_{t+h-1}\left[y_{t+h} \mid y_t\right] \\ &= \widehat{a}^{h-1}\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_t\right) y_t + \frac{\widehat{a}^{h-1} - 1}{\widehat{a} - 1}\widehat{b}\widehat{\sigma}^2_\varepsilon. \end{aligned} \tag{9}$$

In the RW forecasting model, the one-step-ahead or multi-step-ahead point

Figure 4: SB: Forecast density fan chart with a fixed rolling window; the calibration sample size is 50.



forecasts will all be the same, conditional on current information:

$$\begin{aligned} \widetilde{y}_{t+1|t}^{RW} &= E_t\left(y_{t+1} \mid y_t\right) = y_t, \\ \widetilde{y}_{t+h|t}^{RW} &= E_t E_{t+1} \ldots E_{t+h-1}\left[y_{t+h} \mid y_t\right] = y_t. \end{aligned}$$

To evaluate forecasting capacity, we used the conventional RMSFE for both the one-step-ahead and multi-step-ahead forecasting exercises. Both the fixed rolling window and the expanding rolling window forecast are applied. This ad-hoc rolling window approach then allows the estimated coefficients using the AR model to vary over time.

To conduct the fixed rolling window forecast, we choose a window size containing T ($T = 50, 100, 140, 180$) observations for parameter estimations, where T is smaller than the total number of observations N ($N = 201$). Iterated estimation is carried out for periods $N - T$ with a fixed number of observations T, that is, a fixed window size.

Forecasting using an expanding rolling window allows the calibration sample to grow for each calibration period. The initial T observations are used as the first calibration sample. Then, in each iteration with an expanding rolling window, the window expands on adding one more observation.

Table (3) summarizes all RMSFE using the SB, RW, and AR(1) forecasting models, where the (fixed or expanding) rolling window sizes are chosen as $100, 140$, and 180. The SB forecasting model outperforms the RW forecasting model most of the time, providing the smallest RMSFE. From Table (3), as the size of forecast

Table 3: *Root mean squared forecast errors (RMSFE) calculated using the SB, RW, and AR(1) forecasting models. The forecast horizon extends from one-step-ahead forecasts to six-step-ahead forecasts. The fixed rolling window and expanding rolling window approaches are applied, where the window size varies from* 50 *to* 180 *observations. RMSFE values in boldface indicate that SB is favored over RW and AR(1) under the same forecast horizon when the same forecasting method is used.*

	RMSFE	*1-step*	*2-step*	*3-step*	*4-step*	*5-step*	*6-step*
Fix50							
	SB	**0.6554**	**0.6503**	**0.6787**	**0.7261**	0.7853	0.8472
	RW	0.7554	0.7564	0.7564	0.7386	0.7294	0.7097
	AR(1)	0.8238	0.9306	1.0455	1.1398	1.2256	1.2757
Fix100							
	SB	**0.6616**	**0.6546**	**0.6638**	**0.6832**	**0.7039**	0.7342
	RW	0.7564	0.7564	0.7386	0.7294	0.7097	0.7028
	AR(1)	0.7748	0.8254	0.8851	0.9679	1.0507	1.1443
Fix140							
	SB	**0.4595**	**0.4587**	**0.4585**	**0.4573**	**0.4577**	**0.4375**
	RW	0.4793	0.4833	0.4873	0.4863	0.4845	0.4599
	AR(1)	0.5037	0.5767	0.6776	0.7835	0.8937	0.9810
Fix180							
	SB	**0.3882**	**0.3933**	**0.4031**	**0.4106**	0.3581	0.3570
	RW	0.3913	0.3992	0.4096	0.4148	0.3478	0.3351
	AR(1)	0.3596	0.3805	0.4248	0.4984	0.5844	0.6923
Exp100							
	SB	**0.6590**	**0.6439**	**0.6422**	**0.6489**	**0.6554**	**0.6719**
	RW	0.7564	0.7564	0.7386	0.7294	0.7097	0.7028
	AR(1)	0.7623	0.7885	0.8167	0.8644	0.9101	0.9693
Exp140							
	SB	**0.4588**	**0.4580**	**0.4577**	**0.4564**	**0.4566**	**0.4365**
	RW	0.4793	0.4833	0.4873	0.4863	0.4845	0.4599
	AR(1)	0.4958	0.5484	0.6234	0.7003	0.7801	0.8343
Exp180							
	SB	0.3895	0.3957	0.4067	0.4158	0.3662	0.3672
	RW	0.3913	0.3992	0.4096	0.4148	0.3478	0.3351
	AR(1)	0.3556	0.3627	0.3886	0.4402	0.5040	0.5905

Figure 5: SB: Forecast density fan chart with a fixed rolling window; the calibration sample size is 100.

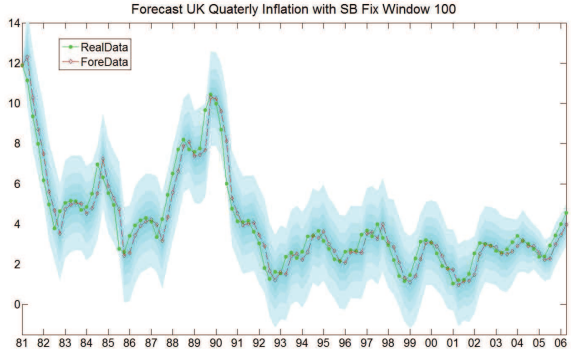


Figure 6: SB: Forecast density fan chart with an expanding rolling window; the calibration sample size is 100.

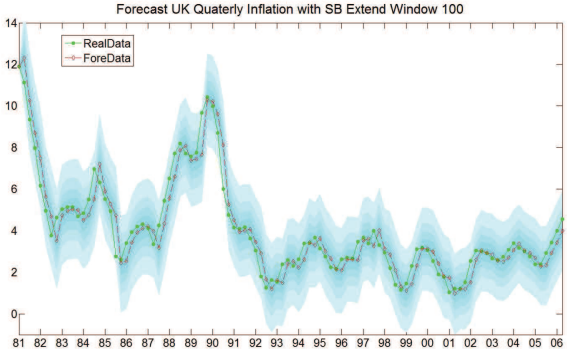


Figure 7: AR(1): Forecast density fan chart with a fixed rolling window; the calibration sample size is 100.

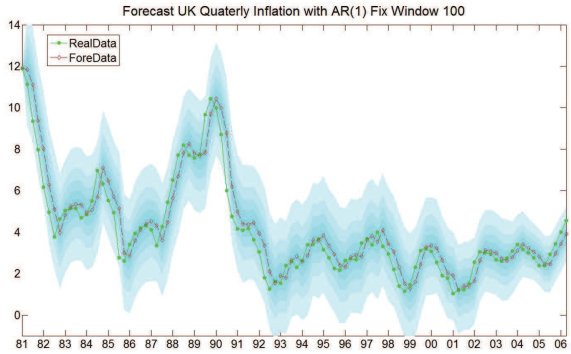
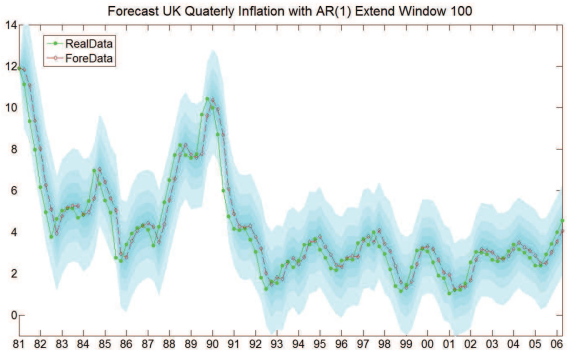


Figure 8: AR(1): Forecast density fan chart with an expanding rolling window; the calibration sample size is 100.



sample $(N-T)$ increases, the forecast errors increase in all the forecasting models.

Using the MCMC draws, it is straightforward to simulate the one-step-ahead and multi-step-ahead predictive distributions. Because the calibrating sample changes as the window rolls, the simulated forecast distributions would, therefore, change over time. We can use fan charts to illustrate the time-varying forecasting distributions, as well as the time-varying forecasting uncertainty. Figures 4– 8 plot the fan charts using the SB and AR(1) forecasting models with different calibrating sample sizes. In Figures 5 and 6, the forecast densities are simulated using a fixed rolling window with the SB forecasting model. Forecast uncertainty is reduced at the end of forecasting samples; that is, the forecasting distribution becomes narrower. This is probably because the estimate $\widehat{\sigma}_{\varepsilon}^2$ becomes smaller

in the last few calibration samples. From Figures 4–8, there are bigger forecast uncertainties; that is, wider forecast distributions, in the mid-80s and mid-90s.

4 Conclusion

This study investigated the first order bilinear model with a single bilinear term using a Bayesian approach. On application to U.K. inflation data from quarter 1-1957 to quarter 1-2007, we found that the SB model is a better fit for examining the underlying process of inflation in the U.K.

The high persistence in the underlying process varies over time. The estimated time-varying autoregressive coefficients are below one most of the time. However, the coefficients are close to one (or exceeding one) if the idiosyncratic shock in the previous period is large and positive. Ignoring the bilinear term b may lead to incorrect inferences, thereby inducing severely distorted forecasts. Motivated by Brunner and Hess (1995) 's views regarding the potential problems in estimating the bilinear model with the MLE, we developed an efficient MCMC sampling algorithm to estimate the SB model where the moment conditions are met.

Overall, the SB model outperforms the RW and AR(1) models in out-of-sample forecasting. In terms of predictive likelihood, the SB model is superior to the RW and AR(1) models for almost all (19 out of 20) of the forecast observations.

Appendix

A Derivation of the Posterior Conditionals

The likelihood function of the SB model can be derived recursively. The simplest SB model is specified as

$$y_t = (a + b\varepsilon_{t-1})\, y_{t-1} + \varepsilon_t, \quad t = 2, \cdots, n,$$

where y_1 is the initial observation and $\varepsilon_1 = 0$. Denote $\mathbf{y} = (y_2, \cdots, y_n)'$ and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$. With $n \geq 2$,

$$\begin{aligned} y_n &= a y_{n-1} + b y_{n-1} \varepsilon_{n-1} + \varepsilon_n, \\ y_{n-1} &= a y_{n-2} + b y_{n-2} \varepsilon_{n-2} + \varepsilon_{n-1}. \end{aligned}$$

Therefore,

$$\varepsilon_{n-1} = y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2},$$

and also

$$\varepsilon_{n-2} = y_{n-2} - a y_{n-3} - b y_{n-3} \varepsilon_{n-3}.$$

If we take the substitute equations recursively,

$$\begin{aligned} y_n &= a y_{n-1} + b y_{n-1} \left(y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2} \right) + \varepsilon_n \\ &= a y_{n-1} + b y_{n-1} y_{n-1} - ab y_{n-1} y_{n-2} - b^2 y_{n-1} y_{n-2} y_{n-2} + ab^2 y_{n-1} y_{n-2} y_{n-3} \\ &\quad + b^3 y_{n-1} y_{n-2} y_{n-3} \varepsilon_{n-3} + \varepsilon_n. \end{aligned} \tag{10}$$

The component with ε_{n-3} can be further substituted. The last component with error term ε_2 then is

$$\begin{aligned} & (-1)^{n-1} \left[b^{n-2} y_{n-1} \cdots y_2 \varepsilon_2 \right] = (-1)^{n-1} \left[b^{n-2} y_{n-1} \cdots y_2 \left(y_2 - a y_1 - b \varepsilon_1 y_1 \right) \right] \\ = \; & (-1)^{n-1} b^{n-2} y_{n-1} \cdots y_2 y_2 + a \left(-b \right)^{n-2} y_{n-1} \cdots y_1 + (-1)^n b^{n-1} y_{n-1} \cdots y_1 \varepsilon_1. \end{aligned}$$

Because we assume $\varepsilon_1 = 0$, equation (10) can be generalized for $t = 3, \cdots, n$ as

$$y_t = a \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right] + \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right] + \varepsilon_t.$$

If we denote

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

$$f_2(t, b) = \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right], \; t \geq 3,$$

equation (10) can be summarized as

$$y_t = af_1(t,b) + f_2(t,b) + \varepsilon_t,$$

where $f_2(t,b) = 0$ with $t = 2$.

The density of elicited prior for a is

$$p(a) \propto \frac{1}{(2\pi \underline{V}_a)^{\frac{1}{2}}} \exp\left\{ -\frac{\left(a - \underline{\mu}_a\right)^2}{2\underline{V}_a} \right\} I_{[|a|<1]}, \tag{11}$$

where $\underline{\mu}_a$ is the prior mean and $\underline{V}_a$ is the prior variance. Because equation (11) does not integrate up to 1, it is preferable to have a prior that allows $p(a)$ to integrate to 1 within the truncated region for model comparison reasons. Hence, a truncated prior of a can be obtained by dividing the improper prior density by a normalizing constant $\Pr(1\,||a|<1|)$. Therefore, the proper prior of a is the following:

$$p(a) = \frac{1}{\Pr(1\,||a|<1|)\,(2\pi \underline{V}_a)^{\frac{1}{2}}} \exp\left\{ -\frac{\left(a - \underline{\mu}_a\right)^2}{2\underline{V}_a} \right\} I_{[|a|<1]}, \tag{12}$$

where the normalizing constant $\Pr(1\,||a|<1|)$ can be calculated as

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

Φ_N is used to indicate the c.d.f of a normal distribution.

The joint prior of h_ε and b conditional on a can be expressed as

$$p(h_\varepsilon, b \mid a) \propto \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon}\Gamma(\underline{\alpha}_\varepsilon)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1\right]}. \tag{13}$$

The conditional prior for $p(h_\varepsilon \mid a)$, then, can be obtained by marginalizing over b in equation (13) as follows:

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

To meet the moment conditions, b is restricted as $|b| < \sqrt{(1-a^2)\,h_\varepsilon}$ in equation(2). The conditional prior of h_ε, $p(h_\varepsilon \mid a)$, can then be obtained as the following:

$$p\left(h_{\varepsilon} \mid a\right) \propto \int p\left(h_{\varepsilon}, b \mid a\right) I_{\left[|b|<\sqrt{(1-a^2)h_{\varepsilon}}\right]} db \propto p\left(h_{\varepsilon}, b \mid a\right) \int_{-\sqrt{(1-a^2)h_{\varepsilon}}}^{\sqrt{(1-a^2)h_{\varepsilon}}} db$$

$$\propto \frac{1}{\underline{\beta}_{\varepsilon}^{\underline{\alpha}_{\varepsilon}} \Gamma\left(\underline{\alpha}_{\varepsilon}\right)} h_{\varepsilon}^{\underline{\alpha}_{\varepsilon}-1} \exp\left(-\frac{h_{\varepsilon}}{\underline{\beta}_{\varepsilon}}\right) \cdot 2\sqrt{\left(1-a^2\right) h_{\varepsilon}}. \tag{14}$$

Then, a proper conditional prior $p\left(b \mid h_{\varepsilon}, a\right)$ for b is

$$p\left(b \mid h_{\varepsilon}, a\right) \propto \frac{1}{2\sqrt{\left(1-a^2\right) h_{\varepsilon}}} I_{\left[|b|<\sqrt{(1-a^2)h_{\varepsilon}}\right]}. \tag{15}$$

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

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

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

$$\overline{\mu}_a = \overline{V}_a \cdot \left[h_{\varepsilon} \sum_{t=2}^{N} f_1\left(t, b\right)\left[y_t - f_2\left(t, b\right)\right] + \frac{\underline{\mu}_a}{\underline{V}_a}\right]. \tag{16}$$

The posterior conditional for h_{ε} follows a truncated gamma distribution $h_{\varepsilon} \sim \Gamma\varepsilon\left(\overline{\alpha}_{\varepsilon}, \overline{\beta}_{\varepsilon}\right) I_{[h_{\varepsilon} > b^2/(1-a^2)]}$, where

$$\overline{\alpha}_{\varepsilon} = \underline{\alpha}_{\varepsilon} + \frac{1}{2}\left(N-1\right),$$

and

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

If the value of $b^2/\left(1-a^2\right)$ is extremely big, h_{ε} has to be sampled from the upper tail of a gamma distribution. For simplicity, we can approximate the tail with an exponential distribution,

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

where $\lambda = b^2/\left(1-a^2\right)$. The inverse *c.d.f* algorithm together with an exponential rejection algorithm can be applied to generate random draws of h_{ε}.

The value of b depends on the values of a and h_ε. The posterior conditional of b is

$$p\left(b|y,a,h_\varepsilon\right) \propto \exp\left\{-\frac{h_\varepsilon}{2}\sum_{t=2}^{N}\left[y_t - af_1\left(t,b\right) - f_2(t,b)\right]^2\right\}\cdot I_{\left[|b|<\sqrt{(1-a^2)h_\varepsilon}\right]}. \quad (17)$$

Because b does not follow a standard distribution from which we can generate random samples, the Griddy–Gibbs sampling method proposed in Ritter and Tanner (1992) is used to get random draws of b. Based on the above posterior conditionals of a, b and h_ε, a Gibbs samplings incorporated with a Griddy–Gibbs sampling algorithm can be carried out in a straightforward way to simulate the posterior distributions of a, b and h_ε.

References

Atkeson, A., Ohanian, L. E., 2001. Are phillips curves useful for forecasting inflation. Federal Reserve Bank of Minneapolis Quarterly Review 25 (1), 2 – 11.

Barnett, A., Mumtaz, H., Theodoridis, K., 2014. Forecasting uk gdp growth and inflation under structural change. a comparison of models with time varying parameters. International Journal of Forecasting 30 (1), 129 – 143.

Bibi, A., Lessak, R., 2009. On stationarity and -mixing of periodic bilinear processes. Statistics & Probability Letters 79 (1), 79 – 87.

Brunner, A. D., Hess, G. D., 1995. Potential problems in estimating bilinear time-series models. Journal of Economic Dynamics and Control 19 (4), 663 – 681.

Byers, J. D., Peel, D. A., 1995. Bilinear quadratic arch and volatility spillovers in inter-war exchange rates. Applied Economics Letters 2 (7), 215 – 219.

Canova, F., 2 2007. G-7 inflation forecasts: Random walk, phillips curve or what else? Macroeconomic Dynamics 11, 1 – 30.

Charemza, W. W., Lifshits, M., Makarova, S., 2005. Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results. Journal of Economic Dynamics and Control 29 (1 - 2), 63 – 96.

Chen, C. W., 1992a. Bayesian analysis of bilinear time series models : a gibbs sampling approach. Communications in Statistics - Theory and Methods 21 (12), 3407 – 3425.

Chen, C. W. S., 1992b. Bayesian inferences and forecasting in bilinear time series models. Communications in Statistics - Theory and Methods 21 (6), 1725 – 1743.

Feng, H., Peng, L., Zhu, F., 2013. Interval estimation for a simple bilinear model. Statistics & Probability Letters 83 (10), 2152 – 2159.

Gelfand, A. E., Dey, D. K., 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society. Series B (Methodological), 501 – 514.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Alexandria, VA: American Statistical Association, 571 – 578.

Geweke, J., 2001. Bayesian econometrics and forecasting. Journal of Econometrics 100 (1), 11 – 15.

Geweke, J., Amisano, G., 2010. Comparing and evaluating bayesian predictive distributions of asset returns. International Journal of Forecasting 26 (2), 216 – 230.

Granger, C., Andersen, A., 1978a. An introduction to bilinear time series models. Banderhoeck and Ruprecht, Gottingen.

Granger, C., Andersen, A., 1978b. On the invertibility of time series models. Stochastic Processes and their Applications 8 (1), 87 – 92.

Hristova, D., 2005. Maximum likelihood estimation of a unit root bilinear model with an application to prices. Studies in Nonlinear Dynamics and Econometrics 9, 1199.

Kim, W. K., Billard, L., Basawa, I. V., 1990. Estimation for the first-order diagonal bilinear time series model. Journal of Time Series Analysis 11 (3), 215 – 229.

Maheu, J. M., Song, Y., 2014. A new structural break model, with an application to canadian inflation forecasting. International Journal of Forecasting 30 (1), 144 – 160.

O'Reilly, G., Whelan, K., Nov. 2005. Has euro-area inflation persistence changed over time? Review of Economics and Statistics 87 (4), 709 – 720.

Osborn, D. R., Sensier, M., 2009. Uk inflation: Persistence, seasonality and monetary policy. Scottish Journal of Political Economy 56 (1), 24 – 44.

Raeburn, E., Lane, J., Peel, D., 1995. Some empirical evidence on the time series properties of four uk asset prices. Economica 63 (251), 405 – 426.

Ritter, C., Tanner, M. A., 1992. Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. Journal of the American Statistical Association 87 (419), 861 – 868.

Sesay, S. A. O., Subba Rao, T., 1988. Yule-walker type difference equations for higher-order moments and cumulants for bilinear time series models. Journal of Time Series Analysis 9 (4), 385 – 401.

Stock, J. H., Watson, M. W., 1999. Forecasting inflation. Journal of Monetary Economics 44 (2), 293 – 335.

Stock, J. H., Watson, M. W., 2007. Why has u.s. inflation become harder to forecast? Journal of Money, Credit and Banking 39, 3 – 33.

Subba Rao, T., Jan. 1981. On the theory of bilinear time series models. Journal of the Royal Statistical Society. Series B (Methodological) 43 (2), 244 – 255.